

**Univerzita Karlova**

**Filozofická fakulta**

Fonetický ústav

## **Bakalářská práce**

Martina Englmaierová

**Percepce průběhů F0 v českých dvojslabičných slovech  
a v analogickém hudebním materiálu**

Perception of F0 courses in Czech disyllabic words and analogous musical  
material

Na tomto místě bych ráda poděkovala doc. PhDr. Janu Volínovi, PhD. za ochotu, velmi přínosné konzultace a příkladné vedení práce. Velmi si vážím podpory mé rodiny a přátel, a času respondentů, kteří se zúčastnili našeho experimentu.

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V ..... dne ..... podpis: .....

## Abstrakt

Práce se zabývá percepcí melodických a temporální změn v hudebních, řečových a smíšených položkách. Sestavili jsme percepční test, který obsahoval česká dvojslabičná slova a jejich hudební analogy. Pomocí algoritmu PSOLA jsme některé z nich manipulovali kroky 0,5 ST, 1 ST a temporálně o 25 ms. Respondenti měli posoudit, zda je dvojice stimulů (hudebních, řečových a smíšených položek) stejná či různá. Test obsahoval celkem 95 položek (75 + 20 opakovaných) rozdělených do tří samostatných bloků a byl zadán osmnácti respondentům. Kromě percepčního testu posluchači vyplňovali dotazník zjišťující míru jejich hudebnosti. Zjistili jsme, že navzdory faktům v uváděné literatuře nejsou manipulační kroky 0,5 ST a 25 ms v řeči postřehnutelné. Nejvyšší úspěšnost vykazovaly položky sestavené ze stejných stimulů, rovněž položky ze stimulů hudebních. Hudebnost se ukázala jako výhoda pro percepci melodie v hudbě i řeči. Na úspěšnost mělo vliv zejména hudební vzdělání respondentů, doba a četnost jejich cvičení na nástroj či zpěv.

**Klíčová slova:** melodie řeči, základní hlasivková frekvence, percepce, právě postřehnutelné rozdíly, zpracování řeči

## **Abstract**

The subject matter of this thesis is perception of melodic and temporal changes in music, speech and mixed items. We assembled a perception test, which contained Czech disyllabic words and their music analogues. Using the PSOLA algorithm we manipulated some of them by 0,5 ST, 1 ST and 25 ms steps. Respondents were requested to decide if the pair of stimuli in the item was the same or different. The test contained 95 items (75 + 20 repeated) divided into three separate blocks and was administered to eighteen people. Besides the perception test listeners filled in the questionnaires, that surveyed the level of their musicality. Despite facts in a mentioned literature we found out that manipulation steps 0,5 ST and 25 ms in speech were not noticeable. The highest success rate showed items assembled from the same stimuli, items from musical stimuli as well. Musicality turned out to be an advantage for the perception of melody in speech and music. The success rate was mainly influenced by music education, the time and frequency of their instrument or vocal practicing.

**Key words:** Melody of speech, fundamental frequency, perception, difference limen, speech processing

# Obsah

<b>Abstrakt</b> .....	4
<b>Abstract</b> .....	5
<b>1. Úvod</b> .....	7
<b>2. Teoretické pozadí problematiky</b> .....	9
2.1 Akustický a percepční aspekt řeči .....	9
2.2 Právě postřehnutelné rozdíly .....	14
2.3 Současné poznatky o vnímání řečové melodie .....	25
<b>3. Metoda</b> .....	36
3.1 Zvukový materiál .....	36
3.2 Testování .....	40
<b>4. Výsledky</b> .....	42
4.1 Úspěšnost .....	42
4.2 Hudebnost .....	50
<b>5. Diskuse</b> .....	55
5.1 Úspěšnost .....	55
5.2 Hudebnost .....	58
5.3 Limity studie a další výzkum .....	59
<b>6. Závěr</b> .....	61
<b>Bibliografie</b> .....	62
<b>Přílohy</b> .....	67

# 1. Úvod

Předkládaná bakalářská práce se zabývá percepcí melodie v řeči a v hudbě. Téma je v dnešní společnosti aktuální a přirozeně přitahuje pozornost, neboť řeč (a pro mnohé i hudba) je pro člověka a jeho komunikaci zásadní. Obě domény sdílejí některé funkce, jiné je naopak odlišují. Lidé se o oba systémy zajímají, hledají v nich propojení a paralely. My jsme se zaměřili na percepci jejich melodické a temporální složky.

Ve druhé kapitole čtenáře seznamujeme s teoretickým pozadím problematiky. V první části této kapitoly uvádíme pohledy autorů dostupné literatury na její akustický a percepční popis. Melodii řeči postupně představujeme skrze vnímání všech jejích složek. Druhá podkapitola se věnuje nejmenším vnímatelným rozdílům. Srovnáváme nejčastěji naměřené hodnoty s výsledky autorů, jejichž názor na věc se liší. Oddíl tak čtenáře postupně provede náhledy na JND (just noticeable difference) frekvence, amplitudy, trvání a okrajově i spektra. Třetí oddíl druhé kapitoly přináší poznatky ze současných studií na toto téma. Uvádíme experimenty autorů Harris a Umeda (1986), Pierrehumbertová (1979), Patel, Iversen a Rosenberg (2006), a Weidema, Roncaglia-Denissem a Honing (2016).

Třetí kapitola se věnuje již našemu experimentu. Podrobně popisujeme jeho metodiku: použitý zvukový materiál i testovací proceduru. Uvádíme, z jakých dat experiment vychází, detailně popisujeme přípravu jednotlivých položek do percepčního testu i jejich manipulace. V podkapitole Testování předkládáme informace o respondentech, zadávání a průběhu samotného percepčního testu a doplňujícího „hudebního dotazníku“.

Ve výsledcích udáváme konkrétní hodnoty úspěšnosti našeho percepčního testu. Zaměřujeme se na úspěšnost dle jednotlivých bloků testu, typů položek i manipulací. Následuje prezentace celkové úspěšnosti jednotlivých respondentů a hodnot jejich konzistentnosti. V položkové analýze probíráme nejvíce a nejméně úspěšné položky. Ve druhé části čtvrté kapitoly se zaměřujeme na hudebnost. Nalezneme zde celkovou úspěšnost respondentů, kteří byli na základě dotazníku označeni jako hudební, resp. nehudební. Opět jejich výsledky porovnáváme dle

jednotlivých typů položek. Následně zjišťujeme výsledky hudebních a nehudebních respondentů postupně podle všech otázek z hudebního dotazníku.

V následující kapitole 5 diskutujeme uvedené výsledky. Na základě našich zjištění vyvozujeme případné obecně platné závěry. Výsledky srovnáváme s uvedenými poznatky z teoretické části. Na závěr nastiňujeme návrhy dalšího výzkumu.



## 2. Teoretické pozadí problematiky

### 2.1 Akustický a percepční aspekt řeči

Jeden z nejočividnějších poznatků o percepci řeči je fakt, že tak činíme relativně rychle a bez námahy. Větší námahy je potřeba pouze při snížené hlasitosti řeči nebo v hlučném prostředí. Obecně se snažíme směřovat pozornost na významy slov, které slyšíme. Ovšem kdybychom se zaměřili na zvuky, vnímali bychom série slov složených z jednotlivých hlásek. Pokud bychom tu samou hlásku slyšeli vícekrát ve více kontextech, pravděpodobně bychom ji vnímali úplně stejně. Avšak jejich reálné akustické vlastnosti by byly zcela jistě odlišné, neboť neexistuje dvojice zvukového signálu a perceptu, které by si stoprocentně odpovídaly (Foss, Hakes 1978: s. 61-62).

Popper a Eccles rozdělují dění světa do tří „podtypů“: Svět 1 označují jako svět hmoty a energie, kam spadá i řečový signál. Svět 2 obsahuje subjektivní zkušenosti, vjemy a emoce, včetně těch, které jsou relevantní pro řeč. Svět 3 je světem informací, včetně symbolických projevů řeči, tj. jazyk. Psychoakustikou rozumí vztah mezi objekty a procesy světů 1 a 2 (Repp, Schouten 1987: s. 273). Je tedy zřejmé, že vztah mezi objektivně popisujícími akustickými veličinami a subjektivním vnímáním není přímočarý, nýbrž je poměrně složitě formován.

Akustický řečový signál lze popisovat z hlediska několika veličin. (Skarnitzl, Šturm, Volín 2015: s. 37). V následujících oddílech se zaměříme na akustický a percepční popis frekvence, amplitudy a trvání.

#### 2.1.1 *Vztah frekvence a výšky*

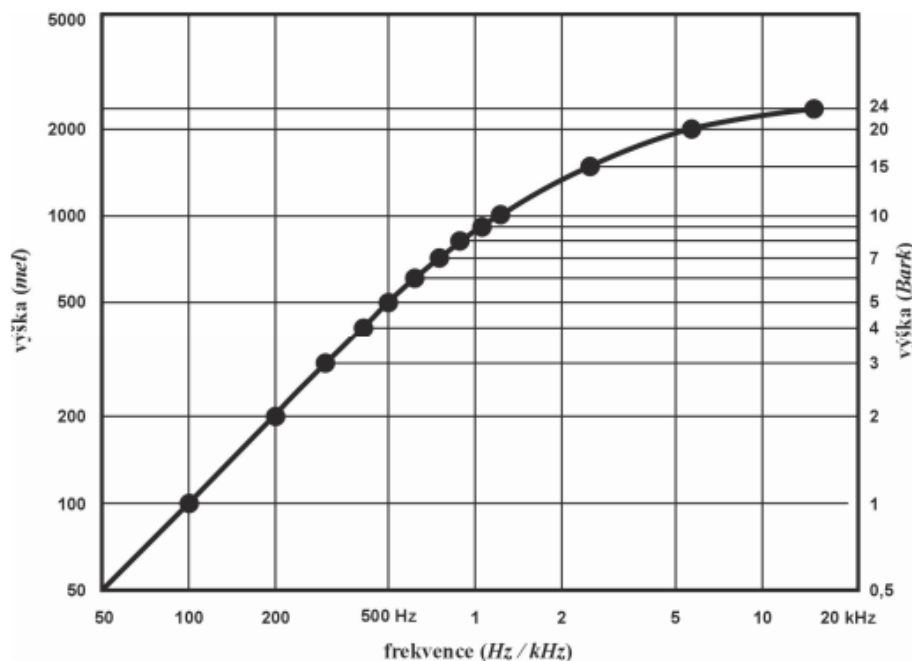
Výška je psychoakustický rys, který v akustickém pojetí řeči představuje rychlost vibrování hlasivek, frekvenci. Přibližné hodnoty základní frekvence kmitání hlasivek se pohybují u mužů kolem 120 Hz, u žen okolo 220 Hz (Cruttenden 1997: s. 3). Ovšem jak jsme již naznačili výše, frekvence není přímým zástupcem výšky.

Jako příklad uvádíme percepci frekvenčních rozdílů. V případě rozdílu 300 Hz a 305 Hz je vzdálenost pěti Hz slyšitelná, v případě 2300 Hz a 2305 Hz již rozdíl nepoznáme. Pokud rozdíl neslyšíme vůbec, je frekvenční vzdálenost menší než nejmenší vnímatelný rozdíl (diference limen; viz oddíl 2.2). Je tedy zřejmé, že lidské ucho je citlivější na frekvenční rozdíly v nižších frekvencích než vyšších. Z tohoto důvodu není vhodné pro subjektivní výšku používat jednotky Hz. Proto bylo navrženo několik jiných vhodnějších jednotek výšky. Až do frekvence 800 Hz, (která pokryje všechny řečové základní frekvence), vnímáme frekvence logaritmicky (poměrově), takže násobky frekvencí (např. 100-200 Hz i 400-800 Hz; v tomto případě dvojnásobek) představují pro lidské ucho přibližně stejné výškové intervaly. Příkladem čistě logaritmické stupnice je stupnice půltónová. Ovšem nad 800 Hz se logaritmické vnímání neuplatní. Pro vyšší frekvence se používá psychoakustických stupnic (viz obrázek 2.1), které vnímané rozdíly objektivizují. Jsou to stupnice melová, barková, nebo erbová. Nejpoužívanější v této oblasti je stupnice erbová, příp. barková (Skarnitzl, Šturm, Volín 2015: s. 40-41).

Melová stupnice byla vytvořena na základě experimentů, ve kterých posluchači subjektivně nastavovali poloviční/dvojnásobnou apod. výšku daného tónu. Z výsledků byla vyvozena stupnice s rozsahem 0 až 2400 mel. Poměry frekvencí v celém rozsahu melové stupnice odpovídají vnímané intervalové vzdálenosti. Platí například 10 Hz = 10 mel, 500 Hz = 500 mel, 1000 Hz = 850 mel, atd.

V rámci psychoakustických stupnic je důležitý pojem *kritické pásmo slyšení* (CB). Jde o frekvenční interval, který vzniká v případě, že dva frekvenčně blízké tóny způsobí dvě vychýlení na bazilární membráně kochley. CB značně ovlivňuje celkovou percepci hlasitosti, která je tvořena součtem amplitud jednotlivých kritických pásem. Na základě CB byla vytvořena stupnice barková a erbová. Obě stupnice vznikly experimentálně pomocí maskování signálního tónu šumem. V případě Barku šlo o rozšiřování pásma šumu. Po překročení určité hranice (1 Bark) se maskování přestalo zvyšovat. Do 500 Hz šířka pásma odpovídá téměř 100 Hz, u vyšších roste s frekvencí (asi 17% střední frekvence). Nevýhodou Barku je hrubé rozlišení v nízkých frekvencích, vznikají rázy mezi tónem a šumem. Podobným způsobem vznikla erbová stupnice, avšak opačně. Pásma šumu se

k signálnímu tónu přibližovala vhodněji z obou stran až do doby, kdy tón začínal být maskován. Vzdálenost pásem šumů v této chvíli odpovídal kritickému pásmu (1 Erb) (Bořil 2016).



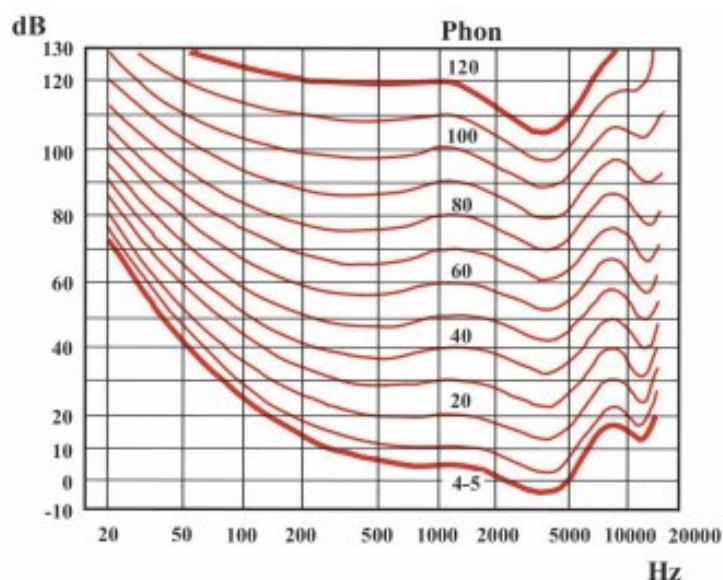
Obrázek 2.1 – Psychoakustické stupnice (Bořil 2016)

Ovšem výška nezávisí pouze na frekvenci. Jistý efekt na celkovou percepci výšky má i hladina zvuku. Pokud uslyšíme čistý tón o frekvenci 200 Hz na hladině zvuku 80 dB, bude se nám zvuk zdát nižší, než kdybychom tutéž frekvenci slyšeli na hladině 40 dB. Naopak vyšší frekvence, např. 6 000 Hz, se při 80 dB bude jevit jako vyšší, a při 40 dB jako nižší (Fastl, Zwicker 2007: s. 113).

### 2.1.2 Vztah intenzity a hlasitosti

S hlasitostí a intenzitou je to ještě o něco komplikovanější. Objektivní síla zvuku se vyjadřuje logaritmicky. Tedy např. zvuk o intenzitě 10 vnímáme jako dvakrát hlasitější než zvuk o intenzitě 1. Stejně tak zvuk o intenzitě 1000 jako dvakrát hlasitější než zvuk o intenzitě 100. Používáme proto pro vyjádření intenzity (resp. hladiny zvuku) logaritmickou jednotku decibel [dB]. Nárůstem 10 dB se

zdvojnásobí vnímaná hlasitost. Tedy např. běžná konverzace o přibližně 50 dB je vnímána čtyřikrát hlasitěji než šeptání o přibližně 30 dB. Avšak přesné určení vnímání hlasitosti určitého zvuku je problematické. Musíme vzít v úvahu jeho trvání, frekvenční složení, aktuální stav organismu mluvčího apod. (Skarnitzl, Šturm, Volín 2015: s. 41-42). Méně používanými jednotkami pro čisté tóny jsou potom hladina hlasitosti (H) a hlasitost (L). Hladina hlasitosti se udává ve fonech [Ph], přičemž při frekvenci 1000 Hz fony odpovídají dB (viz obrázek 2.2). Křivky spojující oblasti stejné hlasitosti se nazývají izofony. Hlasitost (L) s jednotkou [son] se týká celkového vnímání hlasitosti. 1 son odpovídá hlasitosti tónu s frekvencí 1000 Hz a hladinou zvuku 40 dB (Bořil 2016).



Obrázek 2.2 – Izofony (Bořil 2016)

V souvislosti s vnímanou hlasitostí je vhodné uvést rozmezí, v němž je člověk schopen hlasitost zvuku, a zvuk jako takový, vnímat. Spodní hranice sluchu lidského ucha je přibližně 16-20 Hz, horní hranice se udává 20 000 Hz, ovšem tato hranice s přibývajícím věkem klesá (Skarnitzl, Šturm, Volín 2015: s. 42).

Co se týče hladiny zvuku, uvedeme ještě termíny práh slyšení a práh bolesti. Můžeme říci, že práh slyšení závisí na frekvenci. Abychom byli schopni zaznamenat nízké frekvence, např. 100 Hz, musí být hladina tohoto zvuku minimálně 25 dB, zatímco v případě zvuku o 1000 Hz stačí lidskému uchu hladina

zvuku 6 dB. Práh bolesti se pohybuje mezi 120 a 130 dB (Skarnitzl, Šturm, Volín 2015: s. 43).

### ***2.1.3 Vztah trvání a délky***

Percepce trvání není o nic méně komplikovaná než ostatní akustické dimenze. Co se jeví jako odlišné v trvání, bývá často rozdílné ve výšce nebo v jiných parametrech (Fischer-Jorgensen 1961: s. 132). S trváním musí být zacházeno v širším časovém kontextu. Jeho užití je velmi individuální, a to jak na vyšší řečové úrovni, tak na úrovni hláskové (Clark, Yallop, Fletcher 1995: s. 333). Objektivní trvání se od subjektivního příliš neliší v případě, že pracujeme s delšími zvukovými úseky. Ovšem porovnáme-li trvání zvuku či pauzy v řeči, nebo v hudbě, zjistíme, že se vnímaná délka od objektivního trvání může lišit. Podle experimentu Fastla a Zwickera se ve vyšších frekvencích (zde 3,2 kHz) jeví 100ms zvuk, stejně tak pauza, až čtyřikrát delší než je reálný stimul ve skutečnosti (Fastl, Zwicker 2007: s. 265-269). Trvání jednotek řeči je také základem pro vnímání mluvního tempa.

### ***2.1.4 Vztah spektra a barvy***

Barva je definována jako atribut sluchového vjemu, díky kterému je posluchač schopen říci, že dva podobně prezentované zvuky se stejnou hlasitostí a výškou, jsou odlišné (ANSI 1994 v Hardcastle, Laver, Gibbon 2010). Barva je složitým percepčním vjemem, který se nedá jednoduše zaznamenat na číselnou stupnici jako předchozí veličiny. Akustickým termínem pro barvu je spektrum. Jedním z hlavních faktorů barvy je distribuce energie v rámci frekvence (Hardcastle, Laver, Gibbon 2010: s. 469-471). Pokud je spektrální těžiště v nižších frekvencích, zvuk je vnímán jako temnější, pokud je ve vyšších frekvencích, je zvuk vnímán ostřeji (Bořil 2016). Avšak kromě frekvence závisí percepce barvy také na dalších faktorech, například na akustických vlastnostech okolního prostředí i samotného zvuku, na úrovni hláskové také na kvalitách okolních segmentů (Hardcastle, Laver, Gibbon 2010: s. 469-471).

Percepce řeči je do značné míry velice individuální. Jelikož se člověk od člověka liší v mentálních strukturách, nemůžeme vyloučit, ale ani zaručit, že tentýž signál bude vnímán stejně, se shodnou zkušeností a pozorností (Repp, Schouten 1987: s. 11). Záleží také na vizuální podpoře. Příkladem může být McGurkův efekt, který spočívá v tom, že pokud má posluchač k dispozici obrazový záznam osoby pohybující ústy tak, jako kdyby vyslovoval slabiku [ga ga ga...], a zároveň se mu pouští zvukový záznam se slabikou [ba ba ba...], neuslyší [ga] ani [ba], ale [da] (Fernándezová, Smithová-Cairnssová 2014: s. 164).

Z výše popsaného je zřejmé, že mezi subjektivním vnímáním a objektivními realizacemi zvuku je napříč akustickými doménami jistý rozdíl. Ovšem přes tyto nezdá se výrazné odchylky se posluchači na mnoha z nich shodují, což je základní podmínkou fungování řeči.

## 2.2 Právě postřehnutelné rozdíly

Problematika právě postřehnutelných rozdílů je známa pod různými názvy. V češtině se nejčastěji užívá kromě *právě postřehnutelného rozdílu* také *nejmenší vnímatelný rozdíl* nebo *diference limen*. Mezi anglická označení patří *just noticeable difference*, *difference limen*, *least perceptible difference*, *detection threshold*, či *differential threshold*.

V této práci budeme upřednostňovat termín nejmenší vnímatelný rozdíl (JND). O JND mluvíme tehdy, pokud jsme schopni rozlišit zvuky, které jsou téměř totožné vyjma jednoho z aspektů (F0, intenzita, trvání, atd.) (Howard, Angus 2009: s. 137). Postupně podle zmíněných aspektů představíme problematiku a nejčastěji naměřené hodnoty JND.

### Frekvence

JND frekvence značí rozdíl, který jsme schopni postřehnout u zvuků, jež se liší pouze minimálními frekvencí. Začneme-li netypicky, příkladem může být rozeznání dvou současně hraných zvuků. Pokud je náš sluch vystaven dvěma simultánně

hraným čistým tónům, uslyšíme dva různé tóny pouze tehdy, pokud je rozdíl jejich frekvencí větší než kritické pásmo slyšení. V opačném případě vnímáme jeden drsný zvuk, nebo pulzování dle jejich frekvencí (Howard, Angus 2009: s. 258).

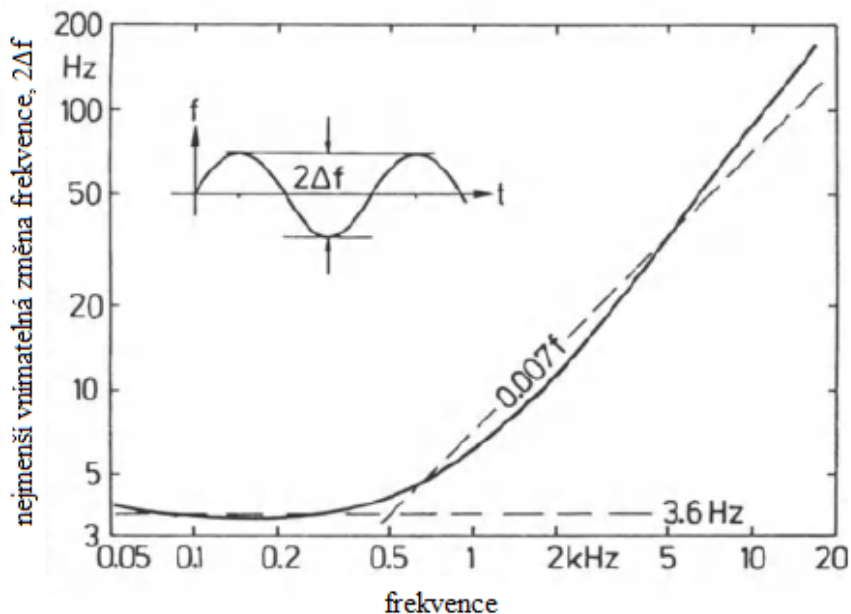
U zvuků, jejichž akustická tlaková vlna je neperiodická, výšku nevnímáme. Ovšem existují neharmonické zvuky, u kterých výšku alespoň omezeně vnímáme. Porovnáme-li percepci výšky [s] v anglickém *sea* a [š] ve slově *shoe*, pravděpodobně se shodneme na tom, že [s] je výš než [š]. Hlávka [š] má totiž spektrální energii vystředěnou v nižších frekvencích a dosahuje maxima kolem 2,5 kHz, zatímco [s] má rovnoměrnější průběh s maximem okolo 5 kHz. Jako příklad neharmonických zvuků v hudbě slouží některé bicí nástroje (Howard, Angus 2009, s. 140-141).

Jak jsme již výše naznačili, hodnoty JND se od autora k autorovi liší, a jeho hodnoty vždy závisí na měřicí metodě a použitých stimulech. A JND frekvence není výjimkou. Zwicker (1957) říká, že JND frekvence sinusového zvuku za stálé hladiny a trvání je asi třicetina kritického pásma napříč rozsahem sluchového pole. Hudebně JND odpovídá přibližně jedné dvanáctině půltónu (Howard, Angus 2009: s. 137).

Fastl a Zwicker jsou názoru, že je důležité rozlišovat mezi nejmenší vnímatelnou změnou (JNV; just noticeable variations) a nejmenším vnímatelným rozdílem (JND; just noticeable difference), což se často v literatuře zaměňuje nebo mísí dohromady. Zatímco JNV se týká postupné změny frekvence, neboli modulace, JND se zjišťuje pomocí dvou stimulů, mezi nimiž je pauza, a to v rozmezí 0,1- 0,2 s (zde použito 200 ms) (Fastl, Zwicker 2007: s. 186).

Průběh JNV můžeme vidět na obrázku 2.3, který znázorňuje JNV frekvence v závislosti na měnící se frekvenci. Odchylka frekvence  $\Delta f$  je definována jako rozdíl mezi původní (nemodulovanou) frekvencí a maximální frekvencí v daném směru. Celková změna frekvence je potom  $2\Delta f$ . Jak můžeme z grafu vyčíst, naše sluchové ústrojí je nejcitlivější na změny frekvence sinusových tónů kolem 4 Hz. V nízkých frekvencích je  $2\Delta f$  poměrně konstantní, a až do přibližně 500 Hz si zachovává hodnotu 3,6 Hz. Na 50 Hz odpovídá  $2\Delta f$  hudebnímu půltónu. Lidský sluch je na změny v sinusových tónech v nízkých frekvencích poněkud méně citlivý než na změny v tónech hudebních, které obsahují mnohem více frekvenčních složek. Tento jev se promítá například v procesu ladění hudebního nástroje, kdy jsou pro naše

ucho vodítkem hlavně vyšší harmonické frekvence. Nad přibližnou hranicí 500 Hz nejmenší vnímatelná změna stoupá téměř úměrně s frekvencí. Má hodnotu přibližně 0,7% frekvence (Fastl, Zwicker 2007: s. 183).



Obrázek 2.3 - Nejmenší vnímatelná změna frekvence v závislosti na frekvenci. Přerušované čáry značí užitečnou aproximaci křivky. Celková změna frekvence je  $2\Delta f$ . Frekvence modulace jsou 4 Hz a hlasitost 60 Ph. (podle Fastl, Zwicker 2007: s. 183).

JND frekvence je podobná JNV, avšak je o 3 jednotky nižší, což je pozoruhodné. Sluchový systém je tedy schopen rozeznat snáze rozdíl frekvencí než její modulaci. Zdá se, že pauza mezi signály (použitá při měření JND) percepce neruší, naopak působí v její prospěch. Posuneme-li výsledky z obrázku 2.3 o 3 jednotky směrem dolů, dostaneme výsledky pro JND frekvence: do frekvence 500 Hz je JND stabilní, a to kolem 1 Hz. Nad 500 Hz je tato hodnota pohyblivá, úměrně roste s frekvencí (přibližně 0,2%  $f$ ). Na hladině zvuku je JND frekvence závislá pouze pod 25 dB. Pod touto hladinou JND frekvence roste v závislosti na ubývajícím hladině zvuku. Frekvenční JND je tedy např. na hladině 5 dB 5 krát větší než na hladině 25 dB (Fastl, Zwicker 2007: s. 186).

Všechny uvedené výsledky JNV i JND jsou vyvozeny na základě více než 200ms stimulů, které již můžeme považovat za poměrně konstantní zvuky. Pro stimuly kratší než 200 ms by se JND frekvence zvýšila (Fastl, Zwicker 2007: s. 186).



t'Hart, Collier a Cohen zdůrazňují, že pro percepci výšky tónu je ve většině případů stěžejní jeho okolí. Určení výšky samostatně stojícího čistého tónu je pro většinu lidí prakticky nemožná. V experimentu Bachema (1937) bylo sedm lidí žádáno označit výšku jednotlivých tónů. Ovšem to je pro běžného člověka více než nesnadný úkol. Chybovali v rozmezí 5-9 půltónů. Pouze jedinci s tzv. absolutním sluchem jsou schopni tento úkol vyřešit, a to s přesností čtvrt půltónu nebo i menší. Pokud přidáme další tón, je pravděpodobné, že posluchač za určitých podmínek rozezná již zmíněný rozdíl až 1 Hz. Uvedené podmínky se týkají frekvence a trvání. Pod 1000 Hz a ve velmi vysokých frekvencích (nad 4 kHz) naše rozlišovací schopnost klesá (t'Hart, Collier, Cohen 1990: s. 27). t'Hart a jeho kolegové uvádějí poněkud odlišné rozpětí hodnot pro nejúspěšnější percepci JND. Tento rozpor s tvrzeními ostatních autorů si vysvětlujeme jako efekt zprůměrování naměřených hodnot.

Závislost na ostatních atributech zmiňují t'Hart, Collier a Cohen v souvislosti s časem. Uvádějí experiment Sergeanta a Harrise (1962), kteří se zabývali časem potřebným pro detekci změny výšky. Z jejich výsledků vyplývá, že během 75 ms je třeba původní frekvenci 1 500 Hz změnit o 30 Hz, tj. 390 Hz/s pro úspěšnou percepci netrénovaných posluchačů. Pro trénované posluchače je třeba pouze 150 Hz/s během 100 ms. S rostoucím trváním se rychlost a potřebná změna frekvence snižuje. Pollack (1968), který experiment Sergeanta a Harrise opakoval s frekvencemi 125, 250, 500 a 1 000 Hz, a s trváním modulace 0,5, 1, 2 a 4 s, přišel s výsledky, podle nichž pokaždé, když je počáteční frekvence o oktávu zvýšena, se práh detekce zvýší o 1,75 Hz/s. To znamená, že percepce je lepší ve vyšších (počátečních) frekvencích. Například pro 125 Hz je potřebná frekvenční změna 2,5 až 3 %, zatímco pro 1 000 Hz pouze 1,7 %. (t'Hart, Collier, Cohen 1990: s. 30).

t'Hart a jeho kolegové dávají JND frekvence také do souvislosti s hlasitostí. Totiž pokles 10 nebo 20 dB v několika desítkách milisekund, což můžeme nalézt u přechodů ze samohlásky na souhlásku, může zcela překrýt až půloktávové změny F<sub>0</sub>. Proto pro výzkum změn F<sub>0</sub> v řeči navrhuje používání celých řečových promluv, a to i v případě, že se v řeči mění jen F<sub>0</sub> (t'Hart, Collier, Cohen 1990: s. 36).

Lehistová (1970) se stejně jako většina autorů přiklání k hodnotě nejmenšího vnímatelného rozdílu 1 Hz. Ovšem setkáme se i s jinými hodnotami. Např. Issachenko a Schadlich (1970) uvádí 5 % (ze 150 Hz, tj. 7,5 Hz)

pro syntetizovanou řeč, Rossi a Chafcouloff (1972) 4 % (ze 195 Hz, tj. 7,8 Hz) pro přirozenou řeč.

Zde ovšem narážíme na problematiku rozdílu mezi percepcí izolovaných čistých tónů a řeči. Zatímco ve většině zmíněných výsledků se jedná o percepci čistých tónů v naprostém tichu a soustředění, percepce řeči na nás klade mnohem náročnější podmínky. To ovšem nutně neznamená, že se percepce výšky v řeči zásadně liší. Největším rozdílem oproti čistému tónu je fakt, že řečový signál je v nejlepších případech kvazi-periodický, a tudíž se s výsledky pro čistý tón budou shodovat spíše vokály (t'Hart, Collier, Cohen 1990: s. 31). Vrátime-li se k nejmenší vnímatelné změně frekvence, zjistíme, že někteří autoři se výsledky svých experimentů blíží hodnotám čistých tónů (t'Hart, Collier, Cohen 1990, Klatt 1973), někteří se s nimi naopak rozcházejí (Rossi 1971, 1978).

Rozdílu mezi percepcí izolovaných čistých tónů a řeči si je vědom také Clark a Yallop, kteří uvádějí nejmenší vnímatelnou změnu frekvence 0,5 % frekvence, a to přibližně do 1 kHz. Podotýkají ale, že v řečovém signálu je JND o něco vyšší, asi 5 %. Velmi dobře rozlišit frekvence jsme schopni přibližně do 500 Hz. To znamená, že si dokážeme vytvořit harmonickou představu vokálů a sonor do 500 Hz a zaznamenat fonologicky relevantní spektrální vrcholy do 3 000 Hz. Frikativní šumy od 3000 Hz do 5 000 Hz pak o něco hruběji (Clark, Yallop 1995: s. 306).

Dle autorů však aktuální hodnota JND závisí také na základní frekvenci, konkrétním mluvčím a jeho řeči (Clark, Yallop 1995: s. 306). Schopnost citlivě rozeznávat výšku je při lidském jednání stěžejní. Celková výška hlasu je pro nás rozhodující v určování pohlaví, věku, i rozlišování mezi jednotlivými mluvčími. Je ale potřeba si uvědomit, že v řeči jsou důležité hlavně malé výškové rozdíly. Podle mírných prozodických nuancí si posluchač tvoří vlastní postoj a (ne)důvěru k mluvčímu, poznává jeho náladu, aktuální rozpoložení, či zájem<sup>1</sup> (Clark, Yallop 1995: s. 323).

Jak je vidět, autoři se často ve výsledcích příliš neshodují, rozdíly však nejsou nijak závratné. Vezmeme-li v úvahu nejnovější výsledky, můžeme shrnout, že JND frekvence sinusového tónu je asi 1 Hz. V řeči se tato hodnota navýší přibližně na 3-5 Hz, a to do frekvence 500 Hz (Fastl, Zwicker 2007), příp. 1 000 Hz (Clark, Yallop 1995). Ve velmi nízkých a velmi vysokých frekvencích se

---

<sup>1</sup> více viz Cooper a Sorensen 1981

JND zvyšuje, nebereme-li v úvahu vliv trvání a hlasitosti. Dnešní fonetika by však kromě těchto přesných čísel ráda znala vliv kontextu na lokální citlivost posluchače. Je tedy potřeba experimentovat s reálnými řečovými stimuly, které jsou přesně lingvisticky specifikovány.

## **Amplituda**

Je složité přesně určit JND síly zvuku, neboť se mění podle charakteru konkrétního zvuku. Nejčastěji se udává hodnota 1 dB. Pokud je ve zvuku přítomno více vyšších harmonických, JND lze snížit na 0,5 dB (Howard, Angus 2009: s. 98). Ovšem v hlučnějším prostředí, kde se odehrává velké množství konverzací, je rozdíl pod 3 dB zaznamatelný pouze s vynaložením značného sluchového úsilí (McShefferty, Whitmer, Akeroyd 2015).

Hodnota JND amplitudy naměřená Clarkem a Yallopem se shoduje s běžně udávanou hodnotou 1 dB, avšak autoři upřesňují podmínky, za nichž tato hodnota platí. JND 1 dB naměřili při vysoké intenzitě a frekvenci 10 kHz. Za mírnějších intenzit a frekvencí nad 10 kHz uvádějí JND 3-4 dB (Clark, Yallop 1995: s. 308). Přesné hodnoty intenzit, za nichž JND platí, autoři nespecifikují.

JND amplitudy závisí stejně jako jiné nejmenší vnímatelné rozdíly na mnoha dalších aspektech. Howard a Angus se zabývali závislostí na hladině zvuku. Podle nich se JND amplitudy pro širokopásmový hluk při hladině zvuku 20-100 dB(SPL) pohybuje od 0,5 dB do 1 dB. Pod 20 dB(SPL) je lidské ucho na změny v amplitudě méně citlivé. Pro čisté sinusové vlny je JND hlasitosti poněkud odlišná. Výrazně se mění dle amplitudy a frekvence vnímaného zvuku. Např. pro 1 kHz tón se amplitudová JND pohybuje v rozmezí od 3 dB při 10 dB(SPL) do 0,3 dB při 80 dB(SPL) (Howard, Angus 2009: s. 97).

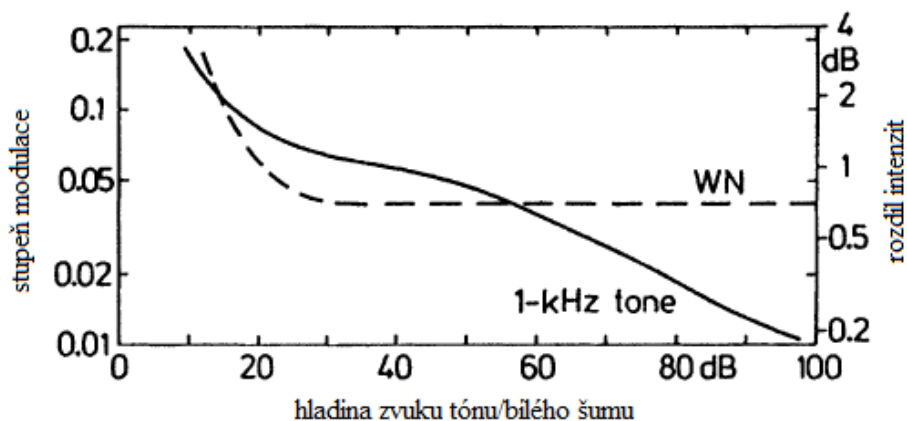
Aktuální vnímaná hlasitost stimulu se značně mění také podle frekvence. Intenzita, při které je daný zvuk již slyšet, se liší o zhruba 70 dB mezi frekvencemi 20 Hz a 15 000 Hz. Proto jsou některé stereo systémy vybaveny schopností zesilovat velmi nízké a vysoké frekvence. V rozsahu 500 Hz až 5 000 Hz, kde se nachází většina informací relevantních pro řeč, disponuje sluchový systém nejnižším prahem vnímatelnosti intenzity, a proto ji v řeči vnímáme vcelku jednotně (Clark, Yallop 1995: s. 308).

Fastl a Zwicker se zabývali měřením nejmenších vnímatelných změn (JNV) v amplitudě, s níž koresponduje rozdíl intenzit tlaku zvuku  $\Delta L$ . Rozdíl intenzit lze vypočítat pomocí stupně modulace amplitudy ( $m$ ), a to rovnicí

$$\Delta L = 10\log(I_{\max}/I_{\min})\text{dB} = 20\log[(1 + m)/(1 - m)]\text{dB} \text{ (Fastl, Zwicker 2007: s. 176)}$$

Graf na obrázku 2.4 zobrazuje vztah mezi stupněm modulace  $m$  a rozdílem intenzit  $\Delta L$ . Plná čára, představující tón o frekvenci 1 kHz, ukazuje, že pro nízké hladiny zvuku je třeba až 20% modulace amplitudy. Na hladinách kolem 40 dB, nejmenší vnímatelná modulace bude již 6%. Se zvyšující se hladinou klesá potřebný stupeň modulace až k 1% na hladině 100 dB. Podobná závislost na hladině zvuku je pozorovatelná u většiny čistých tónů jiných frekvencí (Fastl, Zwicker 2007: s. 176).

Přerušovaná čára pro bílý šum (WN) nemá zcela totožnou trajektorii jako čistý tón. Na nízkých hladinách je pro percepci opět potřeba větší modulace amplitudy, asi 20%. Poté práh zaznamatelné modulace poměrně prudce klesá až k 4% na hladině 30 dB. Na této hodnotě míry modulace bílý šum zůstává až do 100 dB (Fastl, Zwicker 2007: s. 176).



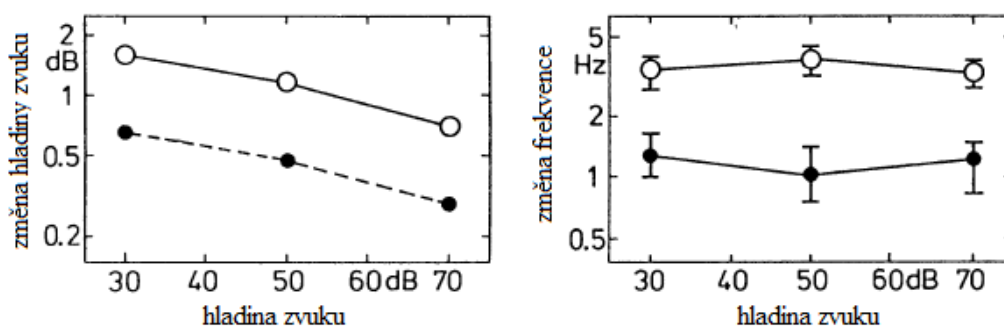
Obrázek 2.4 – Vztah stupně modulace amplitudy a rozdílu intenzit pro 1 kHz tón (plná čára) a bílý šum (WN; přerušovaná čára). Frekvence modulace je 4 Hz (podle Fastl, Zwicker 2007: s. 176).

Zajímavé je, že 6% změny modulace odpovídá změně intenzit 1 dB, což je nejčastější udávaná hodnota ohledně JND hlasitosti. Právě nad hodnotou

odpovídající asi 1 dB se nejmenší vnímatelná změna modulační 1 kHz tónu zvyšuje, a pod 1 dB snižuje (Fastl, Zwicker 2007: s. 177).

Když se podíváme na závislost JNV a JND amplitudy na změně hladiny zvuku a frekvence, zjistíme, že výsledná data pro JNV jsou opět vyšší než data pro JND. Tedy i v případě amplitudy je lidské ucho schopnější rozeznat spíše rozdíl než modulaci.

Z levé části obrázku 2.5 můžeme vyčíst, že JNV se na hladinách mezi 30 a 70 dB pohybuje od 2 dB do přibližně 0,7 dB. Hodnoty JND jsou o přibližně 2,5 dB níže, a to v okolí hodnot 0,7 dB až 0,3 dB pro stejné rozmezí hladiny zvuku. Ovšem toto neplatí pro nejmenší vnímatelné změny frekvence. JNV ani JND amplitudy nevykazují žádnou souvislost s hladinou zvuku (Fastl, Zwicker 2007: s. 180).



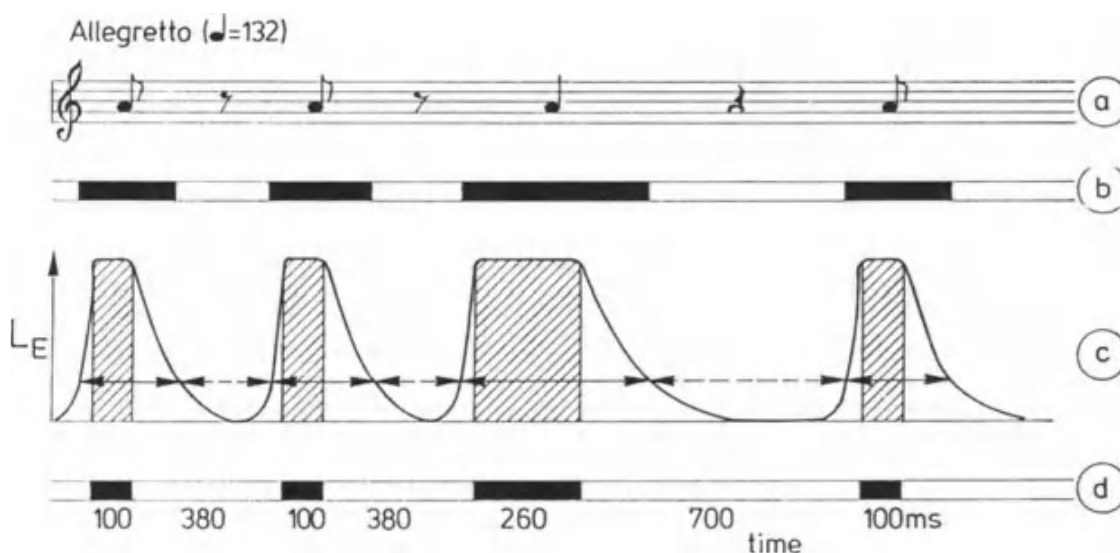
Obrázek 2.5 - Závislost nejmenší vnímatelné změny hlasitosti na změně hladiny zvuku (vlevo) a frekvence (vpravo) pro 1 kHz tón. Prázdné kroužky spojené plnou čarou znázorňují nejmenší vnímatelné změny (JNV), plné tečky spojené přerušovanou čarou značí nejmenší vnímatelné rozdíly (JND). Úsečky na obrázku vpravo značí kvartily, přičemž kolečka jsou dána mediány (podle Fastl, Zwicker 2007: s. 180).

## Trvání

Trvání je určitým ukazatelem rychlosti řeči, avšak s trváním slabik nebo segmentů nemá rychlost řeči přímou souvislost. Je známo, že s rostoucím tempem řeči jsou zachovány pouze akustické aspekty, které jsou nutné pro detekci segmentálních a prozodických vlastností. S chybějícími informacemi se sluchový aparát posluchače vypořádá (Clark, Yallop 1995: s. 322-323).

S tempem řeči neodmyslitelně souvisí rytmus, jehož percepce se napříč jazyky liší. Rytmus je v jazycích jako čeština nebo angličtina vnímán na základě kombinací prominentních a méně prominentních slabik. Jedna z oblastí výzkumu rytmu je založena na p-centrech, tedy percepčních oblastech prominence. Morton, Marcus a Frankish (1976) uvádějí, že řady slabik seřazených podle p-center vykazují větší percepční izochronii než řady s pořadím podle akustických počátků slabik. Studie na toto téma se údajně shodují na tom, že tyto oblasti závisí spíše na trvání a struktuře slabiky, než na jejím přesném segmentálním složení (Clark, Yallop 1995: s. 323).

Autoři Fastl a Zwicker dávají percepci rytmu do souvislosti s hlasitostními vrcholy. Tvrdí, že každé maximum hlasitosti v daném úseku znamená pro percepci rytmickou událost. Ovšem pouze maxima s hodnotou nad 0,43  $N_M$  a rozpětím více než 12%  $N_M$ , kde  $N_M$  představuje hlasitost nejvyššího maxima v rámci daného úseku, např. fráze. Dále takové maximum musí být časově oddělené alespoň 120 ms od dalšího maxima, aby mohlo být rytmickou událostí (Fastl, Zwicker 2007: s. 272-275).



Obrázek 2.6 – Subjektivní vnímání rytmu. a) hudební notace, b) odpovídající trvání rytmických hodnot, c) úroveň percepční excitace odpovídající rytmickému vzoru, d) reálně hrané zvukové události a pauzy. Dvojitě šipky spojené plnou čarou zobrazují subjektivní trvání vzruchu, šipky spojené přerušovaně pak subjektivní trvání pauzy.  $L_E$  značí míru excitace (Fastl, Zwicker 2007: s. 269).

Percepce rytmu je ovlivněna i dalšími akustickými dimenzemi. Ty se navzájem ovlivňují a společně mohou vytvářet percepční změny, které se ve skutečném akustickém signálu neobjevují. Je tedy jasné, že percepce rytmu je do jisté míry neobjektivní. Určité podněty se zdají být delší, silnější, nebo oddělené delší pauzou, než je tomu ve skutečnosti (Skarnitzl, Šturm, Volín 2015: s. 150). Příkladem subjektivního vnímání trvání může být produkce hudebních rytmických hodnot, kterou znázorňuje obrázek 2.6.

Na prvním řádku vidíme původní rytmické hodnoty. Řádek b) vizuálně představuje jeho předpokládanou délku: černé úseky pro noty, bílé pro pauzy. Nicméně v řádku c) vidíme, že hudebníci fyzicky produkují tóny a pauzy s jiným trváním, než je předpokládáno v řádku b). Reálná trvání zvuků a pauz uvádí řádek d) (podle Fastl, Zwicker 2007: s. 273-274). Je tedy zřejmé, že pro vytvoření subjektivně rytmických hodnot je třeba kratších fyzických impulzů. Ovšem musíme podotknout, že autoři zde neuvádějí, za jakých okolností byly výsledky získávány. V běžné praxi by byl kratší fyzický impulz zcela relevantní a možná i záměrný, zvláště pak v uvedeném, poměrně rychlém, tempu. Autoři nezmiňují vliv dozvuku nebo akustické vlastnosti prostředí. Je pravděpodobné, že hudebníci jsou právě kvůli akustickým vlastnostem sálů často na tento způsob navyklí. Musíme také uvést fakt, že není osminová nota jako osminová nota. Velmi záleží na stylu skladby a potřebách dirigenta či interpreta samotného. Za úvahu stojí také otázka, zda by rozdíl mezi řádky b) a d) byl opravdu zanedbatelný. Ve všech případech se jedná o více jak 100ms rozdíl, což je u všech osminových not více jak polovina jejich předpokládaného trvání. Tvzení, že takový rozdíl je nepostřehnutelný, by však vyvracelo níže popsané hodnoty nejmenšího vnímatelného rozdílu.

Konkrétní hodnoty JND trvání uvádí např. Pisoni (1977) Přichází s tím, že v rozpětí 500 Hz až 1 500 Hz jsme schopni rozlišit rozdíl 20 ms. Moore, Peters a Glasberg (1993) se zabývali detekcí pauzy v sinusovém signálu. Nejkratší vnímatelná pauza ve frekvenčním rozpětí 400 – 2 000 Hz je podle nich 6-8 ms. Mimo toto rozpětí až 18 ms. Jistými technikami měření lze okolo frekvence 8 000 Hz dosáhnout JND až 2 ms. Ať už jsou výsledky získány jakoukoliv technikou, je zřejmé, že sluchový aparát je schopen rozeznat pauzy, tedy i počáteční fázi exploziv, dokonce i ve velmi rychlé řeči (Clark, Yallop 1995: s. 308).

Zde přejdeme již ke spektru, ovšem stále v závislosti na čase. Clark a Yallop se zabývali JND spektrálních změn v komplexním signálu. V řeči jde

o postupné změny formantů v průtoku a kodě slabik. Percepční experimenty ukázaly, že rychlé frekvenční změny ve spektru okolo 30 ms jsou temporálně integrovány sluchovým aparátem a zní jako jeden širokopásmový signál. Podle dalších výzkumů je ale očividné, že rychlé změny frekvencí jsou lépe zaznamenatelné, pokud následuje stabilní kontinuální signál. Na hranici vnímání jsou již poměrně rychlé formantové změny teprve okolo 50 ms, které najdeme například u znělých exploziv (Clark, Yallop 1995: s. 308).

## **Spektrum**

Spektrální změny vyvolávají v percepčním pojetí změny barvy. Percepce barvy se nedá zaznamenat na žádnou jednodimenzionální stupnici, hodnotí se subjektivně, a to například číslicemi 1-10 na škále jasný - temný, či pronikavý - tupý. Takových vjemů lze dosáhnout přidáváním a ubíráním energie v různých frekvencích. Důležité pro percepci barvy jsou pravděpodobně vyšší harmonické složky do páté či sedmé. Harmonické vyšší než pátá nebo sedmá, které již sluchově neanalyzujeme, pro barvu již tak důležité nejsou. Mohou však ovlivňovat celkový dojem barvy zvuku jako skupina ležící v určitém kritickém pásmu. Barvy hudebních nástrojů, které mají silné harmonické vyšší než pátá či sedmá, jako například varhany, tenor saxofon, trumpet, housle nebo školený pěvecký hlas, popisujeme jako jasné, brilantní, pronikavé až ostré. Naopak nástroje, které tolik energie ve vyšších harmonických nemají, jako například klarinet, hoboj, flétna, pozoun, lesní roh nebo tuba, člověk hodnotí jako temnější, méně pronikavé až nevýrazné. V druhém případě by bylo možné nástroje rozdělit do skupin podle toho, zda obsahují všechny harmonické do páté či sedmé. Dostali bychom v jedné skupině klarinet, hoboj, flétnu, které podmínku splňují, a v druhé skupině pozoun, lesní roh a tubu, které mají jen několik z nich. Absence sedmé vyšší harmonické (jako v případě flétny) znamená v percepci barvy méně pronikavosti. Např. barva klarinetu je tak specifická díky vlastnímu rozvržení síly v oblasti 6. až 8. harmonické, konkrétně přítomnosti sedmé a slabší síle šesté a osmé harmonické. (Howard, Angus 2009: s. 242-243).



Pro percepci barvy hudebních nástrojů je velmi důležitý úplný začátek zvuku. Rozeznání nástrojů napomáhá již zaznamenání položení smyčce na strunu, nádech před tónem dechového nástroje, počáteční zákmit plátku, tupý úder kladívka uvnitř klavíru, apod. Tato skutečnost je důležitá v syntetické hudbě, kde jsou tyto detaily nezbytné pro dosažení pocitu přirozenosti barvy nástroje (Howard, Angus 2009: s. 242).

Z výše popsaného je očividné, že měřit nejmenší vnímatelné rozdíly ve spektru není jednoduché. Ovšem z nějakého důvodu se na popisu barvy posluchači shodují, proto se někteří autoři snaží zmíněné percepční tendence objektivizovat (např. Frič 2011; Vencovský 2014).

### **2.3 Současné poznatky o vnímání řečové melodie**

Jak jsme již výše uvedli, existuje jistý rozdíl mezi JND frekvence čistých tónů a řečového signálu. Jelikož se velké množství dosavadních studií věnovalo výzkumu čistých tónů, Harris a Umeda (1986) se rozhodli ověřit JND základní frekvence v řeči. Navázali na Flanagana a Saslowa (1958), kteří se zabývali JND konstantní F0 v syntetických vokálech. Trénovaní posluchači byli v tomto experimentu schopni rozeznat rozdíl 0,32 Hz. Autoři experimentu tvrdí, že si jejich posluchači vedli v rozpoznávání rozdílů F0 ve vokálech dokonce o něco lépe než v čistých tónech stejné frekvence i hladiny. Klatt (1973) nesouhlasil s tím, že by se konstantní F0 dostatečně přibližovala řeči, a tak porovnával JND konstantní F0 syntetického vokálu /ε/ ještě s pozvolným náběhem a rychlou změnou F0. JND pro konstantní F0 Klattovi vyšla 0,3 Hz, tedy podobně jako jeho kolegům, pro pozvolný náběh 2 Hz, a pro rychlou změnu (32 Hz/s) vyšla JND 4 Hz. Až poslední případ považuje Klatt za srovnatelný s přirozenou řečí.

Harris a Umeda (1986) na základě dřívějších studií předpokládali, že JND F0 v přirozené řeči bude vyšší než dosud naměřené hodnoty pro neměnnou F0. Ve všech čtyřech experimentech, které Harris a Umeda sestavili, využili reakcí netrénovaných posluchačů. V prvních dvou experimentech posluchači rozpoznávali

změny F0 trvající asi 750 ms (625-825 ms) ve zhruba dvousekundových větách. Druhé dva experimenty obsahovaly kratší věty („The *subject verb*“), asi 600-800 ms, přičemž F0 byla zvýšena či snížena v celé větě.

Pro každý experiment byly užity jiné stimuly. Lišily se v použitých větách, jejich délce, mluvčích (pouze muži), míře a umístění změny F0. Pro změny F0 byl použit LPC vocoder program. Středoškolským či vysokoškolským posluchačům byly v náhodném pořadí předloženy dvojice vět, kde jedna z nich byla s původní konturou F0, a druhá se změněnou F0 (včetně možnosti nulové změny). Respondenti měli říci, zda se druhá věta od první liší. Dvojice vět přicházely po pauze 2,55 s, přičemž mezi jednotlivými větami ve dvojici byl interval 0,55 s.

V prvním experimentu bylo jedním mluvčím čteno pět delších vět s manipulacemi 5 Hz oběma směry. Změny F0 byly umístěny přibližně doprostřed věty tak, aby zahrnovaly přízvučné slovo. Dvaceti posluchačům byly v intervalu jednoho týdne přehrány dva bloky obsahující 180 dvojic, přičemž každý typ dvojice (změna směrem nahoru/dolů a změna první/druhé věty) se mohl opakovat v jednom bloku čtyřikrát. Zatímco směr změny F0 se jevil jako nesignifikantní, výsledné hodnoty JND se u některých vět zásadně lišily. Nejmenší vnímatelný rozdíl byl naměřen necelých 5 Hz, maximální hodnoty dosahovaly 16 Hz. Hodnoty JND byly určovány při 75% úspěšnosti rozpoznání odlišnosti vět.

Ve druhém experimentu se autoři zaměřili na pozici změny F0. V některých větách se totiž začátek nebo konec změněné kontury F0 objevoval na pozicích vokálu, což mohlo vést ke snazšímu určení změny. Manipulovaný úsek umístili tedy tak, aby byl ohraničen explozivami. Druhý experiment obsahoval čtyři věty (namluvené dvěma mluvčími), z nichž se ani jedna nevyskytovala v předchozím experimentu. Bylo užito 3Hz změn F0 pouze se stoupajícím směrem, neboť se neprokázal zásadní rozdíl mezi směry změn. 192 dvojic 2-3sekundových vět bylo předloženo 38 respondentům, přičemž polovina z nich slyšela hlas jednoho mluvčího a druhá polovina druhého mluvčího. Každou dvojici bylo možno slyšet osmkrát. Výsledky JND se pohybovaly u prvního mluvčího od 10 do 15 Hz, (tj. 8-11% frekvence), u druhého mluvčího byly hodnoty vyšší než 15 Hz. Jak se dalo předpokládat, průměrná JND je tentokrát vyšší než v experimentu I. Autoři zde také objevili významný rozdíl mezi mluvčími (nikoli mezi jednotlivými větami).

Dosavadní výsledky autorů této studie ukazují, že JND F0 v řeči jsou 20 krát vyšší než JND konstantní F0 v syntetických vokálech (Flanagan, Saslow 1958), a 2-4 krát větší než JND lineárně pohybující se F0 (Klatt, 1973). Jak bylo předpokládáno, člověk je na změny F0 méně citlivý v přirozené řeči, kde se F0 neustále mění, než na lineární změny F0 syntetických vokálů.

Třetí experiment byl rozdělen na dvě části. První z nich (IIIa) se zaměřovala na otázku, zda jednodušší kratší stimul snižuje hodnotu JND, a druhá část (IIIb) byla zaměřena na vliv různých mluvčích na naměřené hodnoty JND. V první části experimentu III byly použity krátké věty, v nichž se měnila F0 (3 Hz směrem vzhůru) celé věty. Obsahovaly tři jednoslabičná slova (např. „*The train moved.*“). Sedmnáct středoškolských studentů poté hodnotilo 288 párů vět namluvených jedním mluvčím (pro pozdější reference je nutné poznamenat, že jde o mluvčího MH). Byla zjištěna průměrná hodnota JND 4,8 Hz, tedy o mnoho menší než hodnota z experimentů I a II. V experimentu III(b) bylo použito jedné věty z první části experimentu („*The boy talked.*“), neboť se neprokázala výrazná odlišnost jednotlivých vět. Průběh byl stejný jako v předchozí části, ovšem změny F0 se pohybovaly v rozsahu 2 Hz. Pásky namluveny třemi novými mluvčími obsahovaly všechny možné kombinace párů od každého mluvčího, přičemž v každé sadě bylo 144 párů vět. Devatenáct vysokoškolských studentů uslyšelo v rozmezí tří týdnů tři sady zvuků (od každého mluvčího jednu). Ukázalo se, že jeden mluvčí (JW) se výrazně lišil od ostatních dvou (DM a RC). JND mluvčího JW byla 3,25 Hz, mluvčího RC 7 Hz, a mluvčího DM více než 10 Hz. JND mluvčího MH z experimentu III(a) by byla 4 Hz, čímž by se MH přiřadil k signifikantně odlišnému JW. Autoři ovšem neuvádí, že výsledek mluvčího MH pochází z předchozího experimentu, kde byly změny F0 nastaveny na 3 Hz, což mohlo napomoci přesnější percepci. Harris a Umeda podněcují další výzkumy, které by prozkoumaly příčinu, která stojí za zjištěným rozdílem. V diskusi uvádí, že schopnost rozpoznat změny F0 závisí nepřímo úměrně na četnosti jejich výchylek. Tedy čím častěji se F0 v hlase mluvčího mění, tím obtížněji změny F0 detekujeme.

Čtvrtý experiment zkoumal vliv spektrálních úprav signálu (provedených metodou lineární predikce, LPC) na vnímání JND mluvčích z experimentu III. Po úpravě stimulů algoritmem pro spektrální korekci se výrazně snížila hodnota

JND pouze u jednoho mluvčího (RC), proto by dle autorů bylo ukvapené tvrdit, že zvýraznění některých frekvencí má zásadní vliv na hodnocení F0.

S vnímáním nejmenších vnímatelných rozdílů a kontur F0 ovšem souvisí i pojem deklinace, který ovlivňuje pohled na percepci frekvence i dalších akustických atributů. Deklinace je jev dokázaný v mnoha jazycích, jehož výzkum se zaměřuje např. na fyziologické mechanismy, které jsou za postupné snižování F0 zodpovědné, nebo na problematiku využití deklinace v řečové syntéze. Je také dokázáno, že intonační úsek, během něhož probíhá deklinace, se nutně nepřekrývá s nádechovými úseky. Její „nesprávné“ použití může vést k deformaci zamýšleného obsahu věty. Chováním deklinace v angličtině se zabývala Janet Pierrehumbertová (1979). Zkoumala ji v souvislosti s frekvenčním rozsahem věty, amplitudovou konturou a délkou vět. Zjišťovala, do jaké míry s těmito aspekty souvisí.

Jako stimuly byly v jejím experimentu použity věty (jako např. "*The baker made bagels.*"), jejichž slova byla reprezentována slabikami „ma“ (nepřízvučná) nebo „MA“ (přízvučná), (zde tedy „ma MA ma ma MA ma“). Poté, co F0 druhé přízvučné slabiky byla manipulována, středoškolské studenty hodnotili, která přízvučná slabika se jim zdá výš.

V experimentu na frekvenční rozsah byly použity dvě sady, které obsahovaly stejný počet stimulů. V první sadě měl vrchol první přízvučné slabiky frekvenci 151 Hz, tedy o 71 Hz více než počáteční frekvence. Ve druhé sadě měl vrchol první přízvučné slabiky frekvenci 121 Hz, tudíž počáteční vzestup byl pouze 41 Hz. Poté byly druhé přízvučné slabiky postupně manipulovány na 111, 117, 123, 131, 138, 144, 151, 157, 163, 169, 178, 185 a 192 Hz u stimulů s širším frekvenčním rozsahem, a na 96, 102, 107, 112, 117, 121, 126, 131, 136, 142, 147 Hz u stimulů s užším frekvenčním rozsahem. Od každé varianty bylo vytvořeno 10 stimulů, což při devíti respondentech dělalo 90 responsí.

Výsledky první sady stimulů (s širším rozpětím) ukázaly, že aby posluchači slyšeli oba přízvučné vrcholy stejně vysoké, musel být druhý stimul o 9,2 Hz nižší. Nevědomky tedy počítali s deklinací (percepčně normalizovali). Výsledek u druhé sady je překvapivý a obtížně vysvětlitelný. U užšího rozpětí posluchači potřebovali druhý vrchol o 5,6 Hz výše, aby ho slyšeli jako stejně vysoký. Jde nejspíše o iluzi související s kontextově podmíněnými vlastnostmi anglické intonace. Autorka ale neví, s konkrétně jakými.

V amplitudovém experimentu byla použita uměle vytvořená kontura amplitudy. Experimenty amplitudy a délky vět se lišily od předchozího také tím, že kontura F0 druhého vrcholu byla zkopírována podle vrcholu prvního, a poté teprve snižována a zvyšována. Dále v těchto experimentech byla F0 nepřízvučných slabik lemujících přízvučné nastavena na stejnou hodnotu, kvůli krátkodobé tonální paměti. První přízvučný vrchol měl frekvenci 169 Hz, a druhé vrcholy postupně devět hodnot z rozmezí 135-192 Hz. V rámci manipulace bylo pomocí syntézy vytvořeno pět typů stimulů: první přízvučný vrchol o 4 dB, resp. o 2 dB hlasitější než druhý, oba vrcholy přízvučných slabik se stejnou amplitudou, druhý přízvučný vrchol o 4 dB, resp. o 2 dB hlasitější než první vrchol. V prvním experimentu bylo použito těchto stimulů, a posluchači měli opět za úkol rozhodnout se, který vrchol byl vyšší. Z výsledků poskytnutých deseti posluchači je jasné, že amplituda má velký vliv na percepci výšky. S rostoucí amplitudou roste percepčně i výška.

Závislost deklinace na délce věty byla testována ve dvou experimentech. V prvním z nich bylo trvání věty manipulováno podle nepřízvučných slabik. Každá věta obsahovala dvě přízvučné a mezi nimi 1-3 nepřízvučné slabiky. Kontura F0 mezi těmito dvěma přízvučnými slabikami byla vždy stejná, upravená na danou délku věty (0,56 s, 0,80 s, 1,24 s). Druhé přízvučné vrcholy byly manipulovány stejně jako v předchozích experimentech. 13 subjektů hodnotilo 15 variant stimulů. Výsledky ukázaly, že pro percepci totožné výšky obou vrcholů bylo třeba vrcholy druhých přízvučných slabik nastavit o 6,9, 9,2 a 8,4 Hz níže. Rozdíl mezi těmito hodnotami však nebyl statisticky významný, a autorčin závěr je, že deklinace v krátkých až středně dlouhých promluvách není ovlivněna délkou promluvy.

Druhý experiment ohledně délky vět se zabýval závislostí deklinace na počtu přízvučných slabik. Věty v tomto experimentu sestávaly vždy ze sedmi slabik. První sada obsahovala tři nepřízvučné mezi druhou a šestou slabikou přízvučnou (ma MA ma ma ma MA ma), ve zbývajících sadách přibyla ještě jedna přízvučná slabika mezi nimi, obsahovaly tak tři přízvučné slabiky. 18 posluchačů hodnotilo výškovou rozdílnost první a poslední přízvučné slabiky v osmnácti variantách každého stimulu. Autorka počítala s možností, že respondenti budou ve druhých dvou sadách ovlivněni druhou přízvučnou slabikou a nebudou schopni porovnávat opravdu první a třetí přízvučnou slabiku. Provedla tedy opatření, které případné ovlivnění zjišťuje. Ve druhé sadě nastavila druhou přízvučnou slabiku na 149 Hz, a ve třetí sadě na 163 Hz. V každé sadě měl první

přízvučný vrchol 169 Hz a poslední byl manipulován stejně jako v předchozích experimentech. Výsledky ukázaly, že aby posluchač slyšel oba přízvučné vrcholy stejně vysoko, musel by být druhý z nich nastaven v první sadě o 11,1 Hz a ve druhé o 10,9 Hz níže. Avšak rozdíl mezi těmito hodnotami nebyl statisticky významný. Ve třetí sadě s vyšší frekvencí prostřední přízvučné slabiky by druhý přízvučný vrchol musel být o již signifikantních 6,8 Hz nižší, aby byl percepčně stejně vysoko jako vrchol první. Můžeme tedy vyvodit, že s přibývajícím počtem přízvučných slabik se deklinace spíše nezvyšuje. Výsledkům třetí sady se autorka již dále nevěnuje.

Závěrem lze tedy říci, že posluchači normalizují deklinaci při posuzování výškových vrcholů v rámci melodické kontury. Jinými slovy: aby vnímali dva přízvučné vrcholy jako stejně vysoké, druhý z nich musí být ve skutečnosti o něco níže. Pierrehumbertová také potvrdila, že mluvčí očekávají větší deklinaci u vět s větším frekvenčním rozsahem, a že očekávaný sklon deklinace je menší u delších promluv než u kratších. Amplitudový sklon, který je často doprovázen sklonem F0, se ukázal jako součást mentální reprezentace deklinace. Oproti jiným studiím (např. Streeter 1978) dokonce výrazně. Je možné, že vliv amplitudy začíná být takto výrazný až v delších frázích, neboť s těmi právě Pierrehumbertová na rozdíl od jiných studií pracovala.

Novější studie přináší na percepci melodických kontur jiný pohled a vidí ji ve světle dalších vlivů. Patel, Iversen a Rosenberg (2006) přicházejí se studií, která zkoumá vliv národních hudebních vzorů (rytmických a melodických) na prozodii angličtiny a francouzštiny. Prostřednictvím nPVI (normalized pairwise variability index) se rytmickým souvislostem věnovali už Patel a Daniele (2003) na stejných jazycích. nPVI bylo vyvinuto ke zkoumání rytmických rozdílů mezi taktově a slabičně izochronními jazyky, a měří kumulativní míru temporálního kontrastu mezi sousedními prvky ve větě. Melodie ovšem takové měřítko nemá. Zatímco dřívější studie zkoumající rozdílnost řečové melodie angličtiny a francouzštiny pracovaly s celou konturou F0 (Maidment 1976), nebo s výňatky ze začátků vokálů (Lee, Todd 2004), Patel, Iversen a Rosenberg se rozhodli tuto oblast zkoumat pomocí spolehlivé kvantitativní metody. A to jim umožňuje prozogram (Mertens 2004), který konvertuje základní frekvenci věty na diskrétní tóny, jež jsou hlavními vodítky pro naše vnímání. Systém generuje buď statické nebo kinetické tóny.

Řečové stimuly byly tvořeny dvaceti anglickými a dvaceti francouzskými větami z databáze autorů Nazzi, Bertoncini, Mehler (1998), které četly čtyři ženské mluvčí pro každý jazyk. Jsou to krátké věty zpravodajského charakteru, které byly mnohokrát použity v jiných výzkumech (např. Nazzi *et al.* 1998; Ramus *et al.* 1999; Ramus 2002). Obsahují 15-20 slabik a trvají přibližně 3 s. Každá z nich obsahuje asi 16 vokálů. Hudební položky sestávaly z témat 136 anglických a 180 francouzských skladatelů narozených v období 1800-1809, a zesnulých 1990-1999. Hudba konce 20. století je poměrně aktuální a spadá pod vlastenecké období, tudíž se jeví jako vhodná pro tyto účely. Témata vybraných skladatelů byla čerpána z muzikologického slovníku instrumentální hudby *A Dictionary of Musical Themes* (Barlow, Morgenstern 1983). Musela obsahovat alespoň 12 not, které nesměly být narušovány tempovými změnami nebo pauzami, které by ovlivnily měření. Při výzkumu autoři vycházeli z notového záznamu témat. K jednotlivým notám bylo přiřazeno relativní trvání podle typu taktu, přičemž doba se rovnala jedné (např. čtvrt'ová nota ve 4/4 taktu, nebo osminová ve 3/8). Kratší hodnoty byly potom označeny zlomky. Melodie každé věty prošla autokorelačním algoritmem (Boersma 1993). Výchozí parametry byly ponechány, s výjimkou frekvence vyhlazovacího okna, minimální a maximální výšky, které byly nastaveny na 200 Hz, 60 Hz, and 450 Hz ve jmenovaném pořadí.

V rámci rytmu se u hudebních i řečových stimulů měřily dvě hodnoty, nPVI a variační koeficient ( $C_{var}$ )<sup>2</sup> definovaný jako podíl standardní odchylky a střední hodnoty. Ve variabilitě výšky brali v úvahu pouze statické tóny. (Typická věta měla 15 vokálů se statickým tónem a jeden s kinetickým). Kvůli měření variability výšky byla každému rovnému tónu přiřazena vzdálenost (v půltónech) od střední hodnoty výšek všech rovných tónů věty. Poté byl z těchto vzdáleností vypočítán  $C_{var}$ . K získání hodnot pro variabilitu intervalů byly každé dva sousedící statické tóny ohodnoceny intervalem, opět v půltónech. A stejně jako v přechozím případě, byl z těchto intervalových vzdáleností vypočítán  $C_{var}$ . Kvantifikace melodie hudebních témat dosáhli tak, že každou notu opatřili půltónovou vzdáleností od referenčního tónu A440 (440 Hz).

---

<sup>2</sup> Patel, Iversen a Rosenberg používají pro variační koeficient zkratku CV. Z důvodu střetu s důležitou konvencí pro označování slabičných typů budeme pro variační koeficient používat zkratku  $C_{var}$ .

Podle výsledků rytmu mají oba jazyky signifikantně odlišné nPVI, tedy kontraktivitu trvání, i  $C_{var}$ , variabilitu trvání, v řeči. V hudbě se liší v nPVI, nikoli však v  $C_{var}$ . Je tedy vyvoditelné, že hudba odráží kontraktivitu řeči, nikoli však variabilitu. Může se zdát, že vyšší  $C_{var}$  vyžaduje vyšší nPVI, a to hlavně u řeči. Ovšem tato tendence nebyla statisticky prokázána. Výsledky pro melodii řeči ukazují, že angličtina a francouzština se ve variabilitě výšky neliší, ovšem v intervalové variabilitě má francouzština signifikantně nižší hodnoty než angličtina. Jinými slovy přechod od jednoho vokálu k dalšímu je výškově monotónnější u francouzštiny než u angličtiny. Stejně tak u melodie hudby. Variabilita výšky se neliší, ale intervalová variabilita je ve francouzštině signifikantně nižší. Tedy dva sousední tóny mají monotónnější charakter spíše ve francouzské hudbě než v anglické. A to i přes to, že absolutní hodnota průměrného rozpětí intervalů je v obou doménách obou jazyků téměř identická. Je zajímavé, že intervalová variabilita je mnohem výraznější v řeči (0,88 vs. 0,68) než v hudbě (0,76 vs. 0,71). Znamená to tedy, že jazykový materiál se liší více než materiál hudební.

Přepočteme-li hodnoty melodie na melodický interval variability (MIV;  $100 \times C_{var}$  intervalové variability), můžeme ji zobrazit společně s rytmem (nPVI) v tzv. RM prostoru (Rytmus, Melodie). Poté můžeme určit tzv. prozodickou (euklidovskou) vzdálenost (pd; prosodic distance) jazyků, kterou lze definovat jako rozdíl vzdáleností bodů pro průměrné nPVI a MIV jazyků. Pro anglickou a francouzskou řeč lze tuto vzdálenost vyjádřit touto Pythagorovou rovnicí:

$$pd(Es, Fs) = \sqrt{(nPVI_{Es} - nPVI_{Fs})^2 + (MIV_{Es} - MIV_{Fs})^2}$$

(Patel, Iversen, Rosenberg 2006),

kde  $Es$  (English speech) znamená anglický řečový stimulus,  $Fs$  (French speech) francouzský řečový stimulus, obdobně pro hudbu ( $Em$ ,  $Fm$ ). Pd mezi anglickou a francouzskou řečí je 27,7 RM jednotek, a mezi hudebními stimuly 8,5 RM jednotek, což je asi 30% řečového rozdílu. Další zajímavostí spojení RM prostoru je fakt, že uděláme-li přímku mezi výslednými body pro řeč, bude svírat poměrně malý úhel ( $14,2^\circ$ ) s přímkou spojující body pro hudbu. Tedy pohyb od francouzské řeči k anglické je velmi podobný pohybu od francouzské hudby k hudbě anglické.



Podíváme-li se na RM prostor jednotlivých skladatelů, i tam mají obě národnosti svůj vymezený region. Výsledky autorů naznačují, že společné vlastnosti řečového rytmu a melodie ovlivňují styl národní hudby.

Vztahem melodie v řeči a hudbě se zabývali také Weidema, Roncaglia-Denissem a Honing (2016), kteří na rozdíl od autorů předchozí studie, a zároveň všech dosavadních, zkoumali souvislost řeči a hudby na totožných stimulech. Tedy velmi podobně jako my v našem experimentu (viz níže).

Autoři se konkrétně zabývali otázkou, zda je percepční zpracování melodie stejné u řeči i u hudby. K dispozici měli 48 respondentů (z toho 21 mužů) ve věku přibližně 25 let. Pracovali se třemi skupinami respondentů: 16 mandarínských mluvčích, 16 nizozemských hudebníků a 16 nizozemských „nehudebníků“. Těm pak přehrávali vytvořené řečové a hudební stimuly. Řečové stimuly čtené rodilou mandarínskou mluvčí se skládaly ze tří dvouslabičných mandarínských minimálních párů, které se lišily tónem druhé slabiky. Na druhé slabice se mohl objevit stoupající tón (tón 2) a tón klesající (tón 4). Např. *tian1 ming2* („dawn“) vs. *tian1 ming4* („destiny“). Tónové kontinuum druhých rozdílů slabik bylo rozčleněno na 11 stupňů oddělených 0,09 ERB, přičemž každý z nich byl přítomen v lineárně stoupající a klesající kontuře F0. Největší použitá vzdálenost byla 45 Hz (4 ST), nejmenší 4,25 Hz (0,3 ST). Počáteční frekvence prvního stupně byla 175 Hz, a konečná jedenáctého stupně byla 220 Hz. Pomocí PSOLY v Praatu byly stimuly takto upraveny, a následně zaměněny s konturami druhých slabik každého slova. V Praatu byly vytvořeny i hudební protějšky, a to extrahováním melodie z řečových stimulů. Hudební položky se v případě druhé slabiky od těch řečových frekvenčně, amplitudově ani temporálně nelišily. Oba druhy stimulů byly normalizovány na trvání 500 ms a hlasitost 80 dB. Před každým minimálním párem byla navíc umístěna 500 ms výplňková jednotka, takže celá trojice trvala 3000 ms. Jednotlivé trojice byly přehrávány po sekundové pauze. Pro každý z jedenácti stupňů kontury bylo vytvořeno šest položek, tedy participanti měli za úkol vyřešit 132 trojic v obou doménách ( $6 \times 11 \times 2 = 132$ ). Pořadí kontur bylo náhodné, ovšem nesměly se vedle sebe objevit dvě stejné kontury. U každého slova se respondenti museli rozhodnout, zda má druhá slabika konturu klesající, stoupající či rovnou. Poté vyplnili dotazník týkající se jejich hudebního vzdělání (počet roků aktivního hraní na hudební nástroj, počet hodin denního cvičení, počet nástrojů, které ovládají).

Percepční citlivost pro každou skupinu respondentů, typ stimulu (řeč, hudba) a typ kontinua (klesající, stoupající) byla hodnocena pomocí skóre  $d'$ .  $d'$  hodnoty byly získávány pomocí transformovaného z-skóre, které odpovídá „úspěšným“ (H; hit) a „neúspěšným“ (FA; false alarm) pokusům:  $d' = z(H) - z(FA)$ . Průměrné  $d'$  hodnoty byly testovány statistickým testem ANOVA za použití skupiny jako meziskupinového faktoru, a kontinua a typu stimulu jako vnitroskupinového faktoru.

Ve výsledcích se objevily odlišnosti v rámci jednotlivých skupin. U mandarínských mluvčích se silně odlišovaly typy stimulu a kontinua.  $d'$  hodnoty hudebních stimulů byly signifikantně vyšší než u řečových, což značí větší citlivost pro výškové změny v hudbě. Klesající tón byl v hudebních stimulech detekován mnohem lépe, než tón stoupající. Řečové stimuly vykazovaly opačnou tendenci, tedy citlivější percepci stoupajícího tónu. Skupina nizozemských nehudebníků vykazovala také odlišnost v kontinuu, avšak stejnou u obou typů stimulů, a to vyšší citlivost pro klesající tón. Skupina hudebníků si vedla lépe v hudebních stimulech, a také byla v obou doménách citlivější na klesající kontury.

Pro zjištění, zda tzv. pozice kategorické hranice je v řeči a hudbě umístěna stejně, provedli autoři další pozorování v každé skupině a pro každou konturu zvlášť. Zatímco žádná skupina nevykazovala signifikantní rozdíly řeči a hudby pro stoupající tón, klesající tóny byly výrazně odlišné v rámci obou domén u všech tří skupin. V mandarínské skupině se klesající kontury řeči a hudby od sebe signifikantně lišily ve všech jedenácti stupních kromě prvního, s vyšší citlivostí pro hudební stimuly. Ve skupině nizozemských nehudebníků se signifikantní rozdíly mezi doménami u klesajícího tónu objevily pouze u stupňů 1-6, a to s větší citlivostí opět pro hudební stimuly. U hudebníků se percepce řeči a hudby klesajícího tónu lišila pouze ve stupni 1 a 3. Je zajímavé, že pro tuto skupinu je percepce hudby v prvních pěti stupních o mnoho lepší, zatímco ve druhé polovině se hodnoty řeči i hudby už do značné míry překrývají.

Výsledky naznačují, že podobnost zpracování řeči a hudby závisí na skupině, tedy na mateřském jazyce a hudebních kompetencích. Ve všech skupinách se ale prokázala větší citlivost na klesající kontury, a to bez ohledu na typ stimulu. Mandarínští mluvčí se nejvíce lišili z hlediska kontinua a typu stimulu, a to s přesnějšími percepčními schopnostmi v hudbě než v řeči, zvláště

v klesající kontuře. (Nízké hodnocení klesající kontury v řeči může být způsobeno tím, že mandarínský klesající tón je sestaven z krátkého klesání za větší intenzity, než byla použita v tomto experimentu, proto mohla působit mandarínským mluvčím potíže). Nizozemští nehudebníci si vedli nejhůře, a to bez rozdílu typů stimulů. Autoři upozorňují na zajímavý fakt, že vyšší hodnoty nizozemských mluvčích pro klesající tón mohou být způsobeny tím, že holandské intonační vzory a ohraničení fráze jsou detekovány právě pomocí klesající kontury. (Např. u francouzštiny bychom čekali vyšší hodnoty naopak pro stoupající tón, neboť se jejich intonační vzory chovají opačně. Z tohoto důvodu autoři navrhuji, aby další výzkumy obsahovaly různé typy jazyků v tomto ohledu). Hudebníci vykazali nejlepší přesnost kategorizace klesajícího i stoupajícího tónu v obou doménách. V obou oblastech využili sluchového tréninku, který jim hudba nabízí. Platí tedy tvrzení, že dostatečná zkušenost s hudbou nebo tónovým jazykem ovlivňuje způsob vnímání výšky.

### 3. Metoda

Studie této práce vychází z předpokladu, že melodické a temporální aspekty se v hudbě a řeči percepčně liší. Stejně tak jako Weidema, Roncaglia-Denissem a Honing (2016) testujeme naši hypotézu pomocí totožných kontur F0 v řečových i hudebních stimulech, což považujeme za zásadní z pohledu objektivity měření.

Pro percepční test byly vytvořeny tři sady stimulů sestávajících z řečových a hudebních typů. Z těch bylo sestaveno 75 dvojic: řeč-řeč, hudba-hudba, řeč-hudba, z nichž 24 sestávalo ze shodných stimulů, a 51 dvojic z odlišných stimulů. Úkolem posluchače bylo rozpoznat, zda jsou stimuly v poloze ve dvojici rozdílné, či nikoliv. Nejprve bylo provedeno zkušební testování na malém počtu respondentů, poté byl sestaven percepční test o 95 položkách (24 + 51 + 20 náhodně opakovaných). Více o testovací proceduře níže. Test byl předložen osmnácti respondentům.

#### 3.1 Zvukový materiál

Řečové stimuly byly vybrány z Pražského fonetického korpusu, konkrétně z části Český rozhlas. Nahrávky zpravodajského typu jsou čtené profesionálními mluvčími, jde o hlasatele národních rozhlasových stanic Český rozhlas 1 a 2. Pro percepční test byla vybrána slova ze souvislého čteného textu šesti mluvčích: čtyř žen a dvou mužů. Mluvčí jsme volili podle kvality hlasového projevu a částečně také obsahu zprávy, neboť nás zajímala pouze dvojslabičná slova s vokály /a/, /a:/, /e/, /e:/. Tyto dva (resp. čtyři) vokály byly vybrány z důvodu jejich relativní blízkosti a frekventovanosti v českých dvouslabičných slovech. Zároveň jsme chtěli zpřesnit výsledky studie užším výběrem materiálu, a vyvarovat se tak případného zkreslení způsobeného rozdílnou formantovou povahou českých vokálů. Na slova byla kladena další kritéria v podobě omezení přetury a kody slabik. Hledali jsme dvouslabičná slova obsahující slabiky s jednoduchou přeturou a ideálně nulovou kodou. Samozřejmě jsme se také vyhýbali slovům s jednoslabičnou předložkou. Ovšem nalézt kvalitně vyslovená slova s těmito kritérii nebylo snadné ani u profesionálních mluvčích. V konečném výběru slov máme tedy i slabiky s jednoduchou kodou a se shluky maximálně dvou konsonantů

v préúře. Zároveň jsme se snažili slova nevybírat z úplných začátků nebo konců promluvových úseků z důvodu nápadných temporálních změn či fonetických redukcí.

Vybraná disylaba (viz tabulka 3.1) jsme v programu Praat (Boersma, Weenink 2014) vyextrahovali s ohledem na „měkké“ začátky a konce zvuků, a v programu Audacity vložili před začátek a za konec každého slova 1 s ticha. Ve stejném programu jsme u každého slova vygenerovali F0 a ručně opravili chyby. Následně jsme pomocí skriptu změřili průměrné hodnoty F0 z prostředních třetin vokálů a zaznamenali je v půltónech (ST). Ty jsme pak od sebe odečetli, a tak získali interval mezi prvním a druhým vokálem každého slova. Hodnoty v hertzech jsme převedli na nejbližší tón v temperovaném ladění. Tyto tóny jsme nahráli na violu v tempu co nejpodobnějším řečovému stimulům. Řečové stimuly byly vždy nejprve přehrány jako předloha, a následně několikrát napodobeny hrou. Přesná výška tónu byla při nahrávání kontrolována pomocí elektronické ladičky AROMA AMT-550B. Zvuky byly nahrány v tiché místnosti bez ozvěny na diktafon ZOOM H4n. Jelikož nám rozsah violy nedovoloval nahrát některé (hlavně mužské) hlasy, příslušné tóny (týkajících se tří slov) jsme nahráli o čistou kvintu výše, a poté konturu pomocí algoritmu PSOLA (Pitch-Synchronous Overlap and Add) v Praatu snížili na požadovanou výšku. Na základě toho, že náš výsledný vzorek obsahoval mnohonásobně větší počet slov od mluvčích ženských, jsme se rozhodli, že položky mužské použijeme do zácvičných položek (vytvořených ze tří slov). Ostrý test tedy obsahuje pouze hlasy ženské (dvacet slov).

Z nahraných hudebních zvuků jsme v Praatu vyextrahovali nejpodarenější pokusy. Ty jsme poté v programu Adobe Audition<sup>3</sup> převedli na monofonní, vzorkovací frekvenci nastavili na 32 kHz, bitovou hloubku na 16 bitů, amplitudu jsme normalizovali na 85%, a opravili polohu nuly v digitálním signálu. Stejně jako u položek řečových jsme hudební zvuky opatřili 1s ticha před a za signálem. Když byly hudební zvuky ve stejné podobě jako stimuly řečové, připodobnili jsme je i temporálně, a to na desítky milisekund. První (resp. druhý) vokál ve slově měl tedy stejné trvání jako první (resp. druhý) tón v hudebním stimulu. Dostali jsme tak položky se shodným celkovým trváním i s trváním jednotlivých složek. Poté jsme

---

<sup>3</sup>zkušební verze na 30 dní volně ke stažení na [http://www.stahuj.centrum.cz/multimedia/mp3\\_a\\_audio/tvorba/adobe-audition/](http://www.stahuj.centrum.cz/multimedia/mp3_a_audio/tvorba/adobe-audition/)

opět pomocí algoritmu PSOLA vygenerovali melodické kontury obou typů stimulů, a tu hudební připodobnili kontuře řečové. Přímé přenesení řečové kontury do hudební jsme zavrhnuli z důvodu nepřirozeně znějícího výsledku. Jako efektivnější se jevílo hudební konturu mírně upravit dle řečové, a poté ji vložit do řečových stimulů namísto té původní. Tím jsme docílili přirozenosti u hudebních i řečových stimulů, a zároveň naprosté identity na úrovni melodie. Jakmile byly hotovy oba typy stimulů, sestavili jsme první typ položek - nemanipulované, tedy typ S (stejný). Řečové stimuly a jejich hudební protějšky osmi vybraných slov (subjektivně vyhodnocených jako zřetelně stejných v rámci řečových vs. hudebních stimulů) jsme tedy v Adobe Audition vložili za sebe v pořadí řeč – hudba (RH). Obdobně jsme vytvořili položky stejného typu (S) řeč-řeč (RR), hudba-hudba (HH). Vzniklo tak 24 položek: tři varianty (RR, HH, RH) od každého z osmi slov.

Pro manipulované položky jsme nejprve museli zjistit melodické kroky, které v položkách použijeme. Vytvořili jsme tedy několik pilotních položek se změnou 0,5 ST, 1 ST a 1,5 ST oběma směry, a to na první či druhé slabice. Ukázalo se, že krok 1,5 ST je příliš zřetelný, a 0,5 ST naopak odhalen jen v některých případech (či pouze sluchově nadanějšími respondenty). Častěji byl krok 0,5 ST rozpoznán, pokud byl na druhé slabice. Druhé slabiky měly také oproti slabice první ve většině případů rovnější konturu. Na základě informací z pilotáže byly tedy do manipulovaných položek použity kroky 0,5 ST a 1 ST oběma směry na druhých slabikách. Temporální změny jsme arbitrárně stanovili na 25ms. Ze zbývajících dvanácti slov bylo tedy vytvořeno 36 R (různý) manipulovaných položek: u každého typu položky (RR, HH, RH) byla druhá slabika zmanipulována u tří slov krokem +1 ST, u dalších tří slov -1 ST, u dalších tří +0,5 ST a u posledních tří -0,5 ST. Ke každému typu položek (RR-S, HH-S, RH-S, RR+1, RR-1, RR+0,5, RR-0,5, HH+1, HH-1, HH+0,5, HH-0,5, RH+1, RH-1, RH+0,5, RH-0,5) jsme vytvořili její ekvivalent s temporální změnou: vložili jsme 25 ms ticha mezi vokály (resp. tóny). Pro temporální změny jsme cíleně volili položky z explozivou (příp. frikativou) mezi vokály z důvodu přirozenějšího nastavení slova.

Vzniklo tak 75 položek (25 RR, 25 HH, 25 RH; z toho 24 typu stejný a 51 melodicky či temporálně manipulovaný), z nichž 20 náhodně vybraných jsme použili dvakrát. Těchto 95 položek jsme pseudonáhodně seřadili tak, aby

nenásledovaly položky utvořené ze stejného slova, položky stejného typu (RR, HH, RH), stejné manipulace (+1/-1/+0,5/-0,5), položky typu S, a temporálně manipulované položky. Vzniklé pořadí bylo rozděleno do tří bloků (po 32+32+31 položkách). Před každou položku byla opět v Adobe Audition vložena 5s desenzitizační pasáž včetně širokopásmového šumu a pípnutí. Po 5-6 položkách jsme střídali sedm různých desenzitizačních zvuků. Po skončení stimulu následovalo 2,5 s ticha. Časové schéma jedné položky bylo následovné: 5 s desenzitizační pasáž – 1,3 s ticho – stimulus - 2,5 s ticho. Respondentům byl test zadáván v rotujícím pořadí jednotlivých bloků. Před ostrým testem bylo využito šesti zácvičných položek mužských mluvčí.

Zácvik	Test
vácha	kraje
naše	rána
zase	večer
	těla
	žena
	váha
	země
	jaké
	praha
	řece
	řeka
	praha
	jedna
	lépe
	chápe
	také
	řada
	dále
	která
	vláda

*Tabulka 3.1 - Seznam použitých dvouslabičných slov.*

## 3.2 Testování

Percepční test byl předložen osmnácti mladým lidem ve věku 21-31 let (průměrný věk = 24, medián = 23). Výběr respondentů nebyl omezen konkrétními požadavky, snažili jsme se pouze o přibližnou věkovou konzistentnost a zastoupení hudebně vzdělaných i nevzdělaných jedinců. Z většiny se jednalo o studenty jazykových oborů, rodilé mluvčí češtiny. Test byl zadáván ve zvukově izolovaném studiu Fonetického ústavu FF UK, či ve velmi tiché místnosti, pomocí notebooku nebo počítače a kvalitních sluchátek. Respondenti byli testováni individuálně, příp. po malých skupinách (max. pět subjektů v dostatečných prostorových rozestupech). Celkové testování trvalo 30-40 min.

Před spuštěním testu byly každému respondentovi podány instrukce, po nichž se mohl dotázat na nejasnosti. Posluchači byli upozorněni na to, že uslyší dvojice zvuků, u kterých se musí rozhodnout a zaškrtnout do záznamového archu (viz Příloha č. 1), zda byly stejné či různé. Při pochybách jim byla k dispozici doplňovací možnost „hádám“. Zadavatelka testu respondentům popsala, jaké typy zvuků uslyší, vč. desenzitizačních pasáží, a upozornila je, že se mají oprostít od odlišných typů položek a zaměřit na melodické a temporální změny. Poté následovala zácvičná část. Pokud měl respondent další stěžejní otázky, které zcela měnily podstatu úkolu, byl zácvik po jejich zodpovězení spuštěn znovu. Po skončení zácviku a ujasnění doplňujících otázek byly postupně spuštěny jednotlivé bloky testu. Mezi každým blokem byli respondenti dotázáni, zda je jim všechno jasné a jestli je vše v pořádku. Ukončení testu následoval krátký dotazník ohledně hudebního vzdělání a aktivity respondenta (viz tabulka 3.2). Záznamový arch a dotazník každého respondenta byl opatřen kódem, který obsahoval pořadí a pohlaví respondenta.



<p><b>Věk:</b></p> <p><b>1) Hudební vzdělání:</b>  ZUŠ KONZERVATOŘ VŠ/VOŠ SOUKROMÉ LEKCE SAMOUK ŽÁDNÉ</p> <p><b>2) Působili jste / působíte v nějakém hudebním tělese?</b> (<i>sbor, trio, kapela, orchestr...</i>)</p> <p><b>3) Jak dlouho jste hráli / hraje na hudební nástroj (resp. zpíváte)?</b> (<i>údaj v letech</i>)</p> <p><b>4) Jak často jste cvičili / cvičíte na hudební nástroj (resp. zpěv)?</b> (<i>např. 3x týdně, každý den, jednou za měsíc, sporadicky</i>)</p> <p><b>5) Jaký máte vztah k tanci?</b> (<i>vyhýbám se mu / nezajímá mě / rád tancuji, když můžu / tancuji amatérsky / profesionálně ...</i>)</p>
--

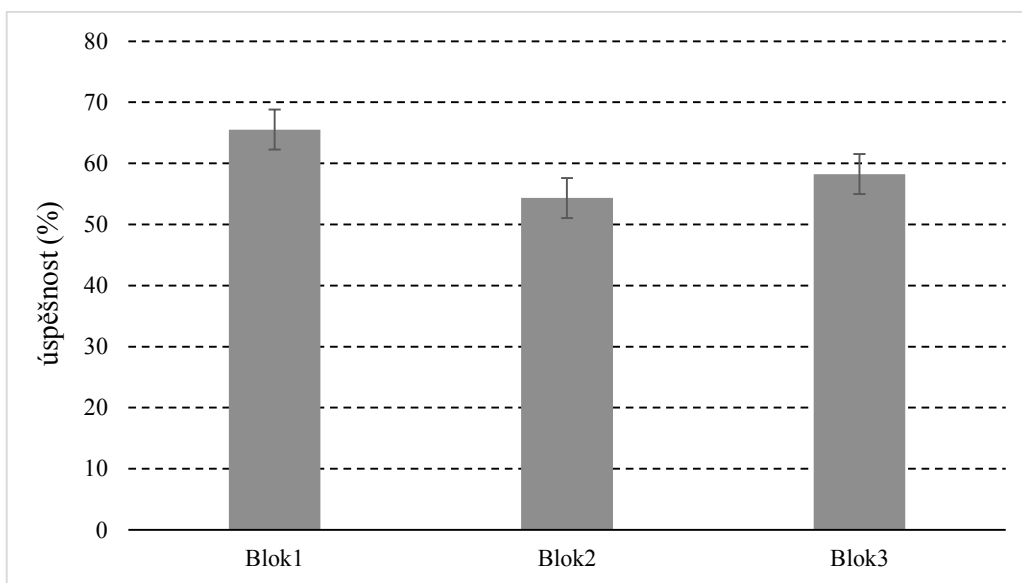
Tabulka 3.2 – Hudební dotazník

Odpovědi na jednotlivé otázky dotazníku byly ohodnoceny 0 - 3 body. Pouze ve druhé binární otázce činila kladná odpověď 2,5 a záporná 0,5. Maximálně bylo tedy možné získat 14,5 bodů. Podrobné hodnocení odpovědí je uvedené v Příloze č. 2.

## 4. Výsledky

### 4.1 Úspěšnost

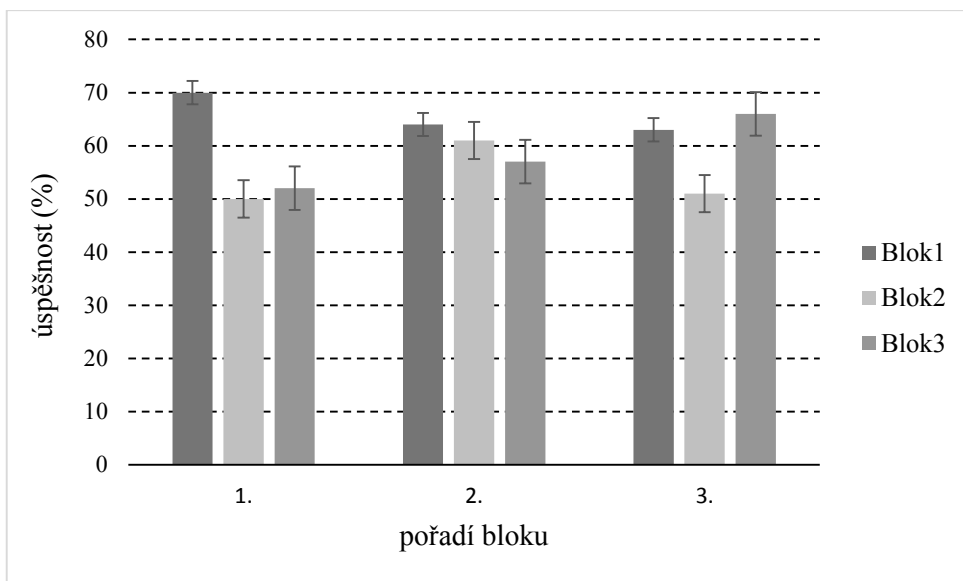
Průměrná úspěšnost percepčního testu dosahuje na 59,4%, přičemž ne každý blok přispěl k celkové úspěšnosti stejně. V grafu na obrázku č. 4.1 vidíme podíly jednotlivých bloků.



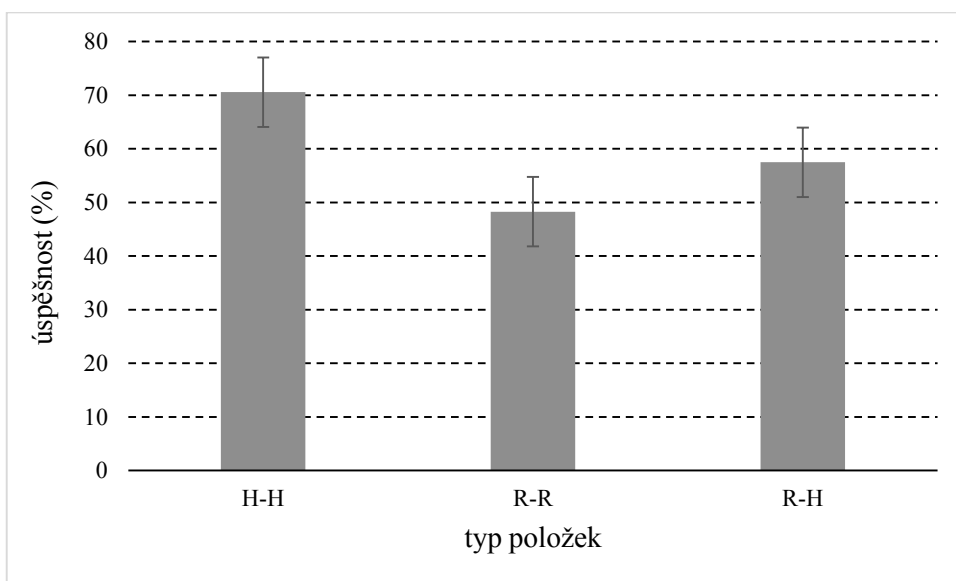
Obrázek 4.1 – Příspevek jednotlivých bloků k celkové úspěšnosti

Blok č. 1 vykazuje nejvyšší úspěšnost (66%). Ovšem podíváme-li se do dat hlouběji, zjistíme, že prokazatelně nejúspěšnější je, pouze pokud byl zadáván jako první v pořadí. Graf na obrázku 4.2 znázorňuje úspěšnost jednotlivých bloků podle pořadí, v jakém byly zadávány. Lze si všimnout, že blok č. 2 má vždy nižší úspěšnost než blok č. 1, a to dokonce kromě druhého pořadí významně. Blok č. 3 je v zadávaném pořadí prvním a druhém také méně úspěšný než blok č. 1, ovšem v pořadí třetím je nejúspěšnější ze všech bloků.

Úspěšnost podle typů položek ukazuje graf na obrázku č. 4.3. Hudební položky vykazují statisticky významně nejvyšší úspěšnost (71%), následují smíšené položky (58%), a nejméně úspěšné byly položky řečové (48%).

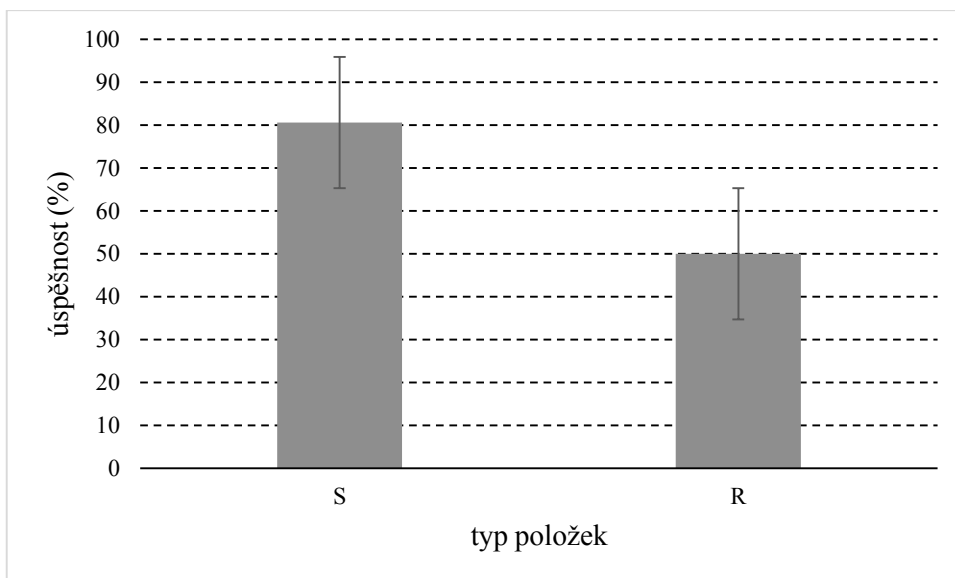


Obrázek 4.2 – Úspěšnost bloků v závislosti na pořadí jejich zadávání.



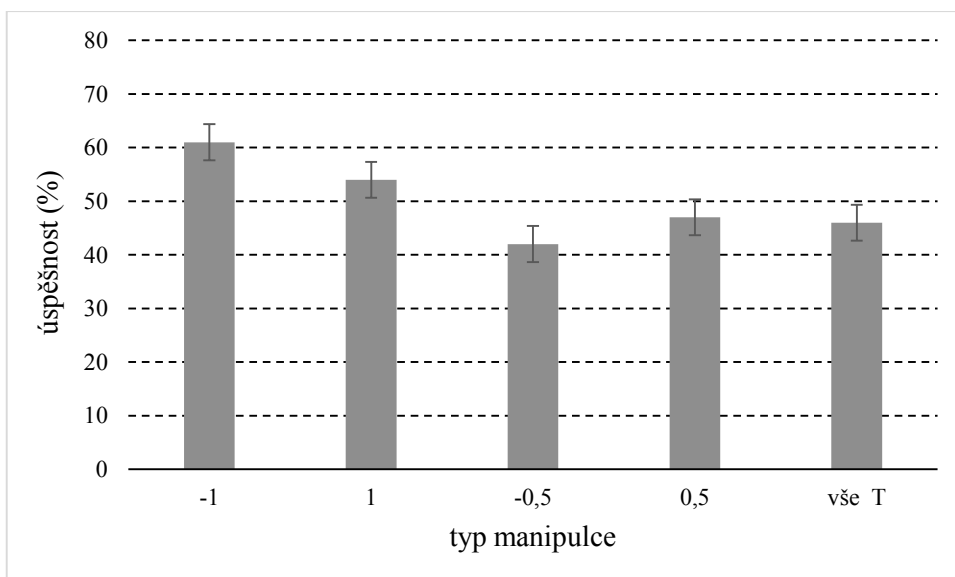
Obrázek 4.3 - Úspěšnost podle typu položek. H-H (hudební položky), R-R (řečové položky), R-H (smíšené položky).

Zajímavé výsledky také uvidíme, podíváme-li se zvláště na úspěšnosti položek S (stejná) a R (různá). V grafu na obrázku 4.4 vidíme, že položky, v nichž byl jeden stimul z dvojice manipulován (R), mají přesně 50% úspěšnost. Položky, které obsahovaly stejné stimuly (S), mají úspěšnost zřetelně vyšší (81%).



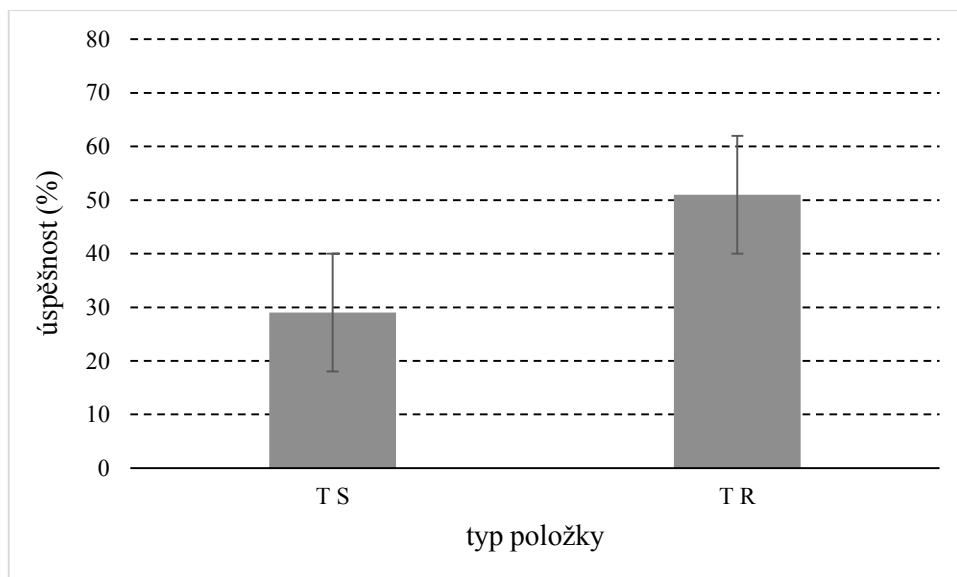
Obrázek 4.4 – Úspěšnost v položkách S (stejně) a R (různé).

Vezmeme-li v úvahu úspěšnost položek podle jednotlivých manipulací, zjistíme, že manipulované položky s krokem 1 ST mají větší úspěšnost než položky s krokem 0,5 ST a dokonce i než všechny temporálně manipulované položky (viz obrázek 4.5).

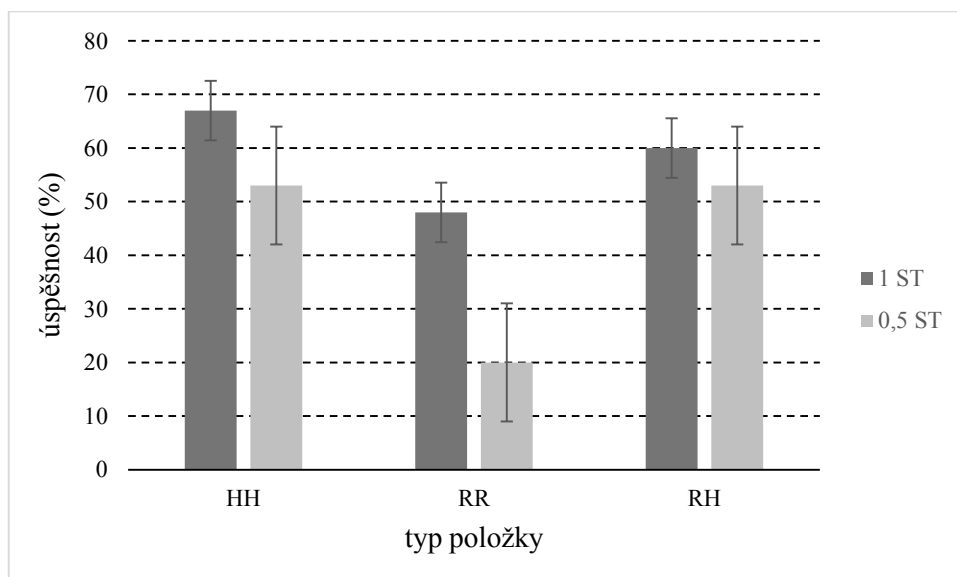


Obrázek 4.5 – Úspěšnost v položkách podle manipulace: -1 ST, +1 ST, -0,5 ST, +0,5 ST, a všechny položky, v nichž byla temporální manipulace (vše T).

V souvislosti s temporální manipulací bychom ještě rádi uvedli (viz obrázek 4.6), že úspěšnost byla významně větší u položek, kde temporální manipulace doplňovala manipulaci melodickou (51%). V položkách, kde tvořila jedinou změnu, byla úspěšnost nižší (29%).



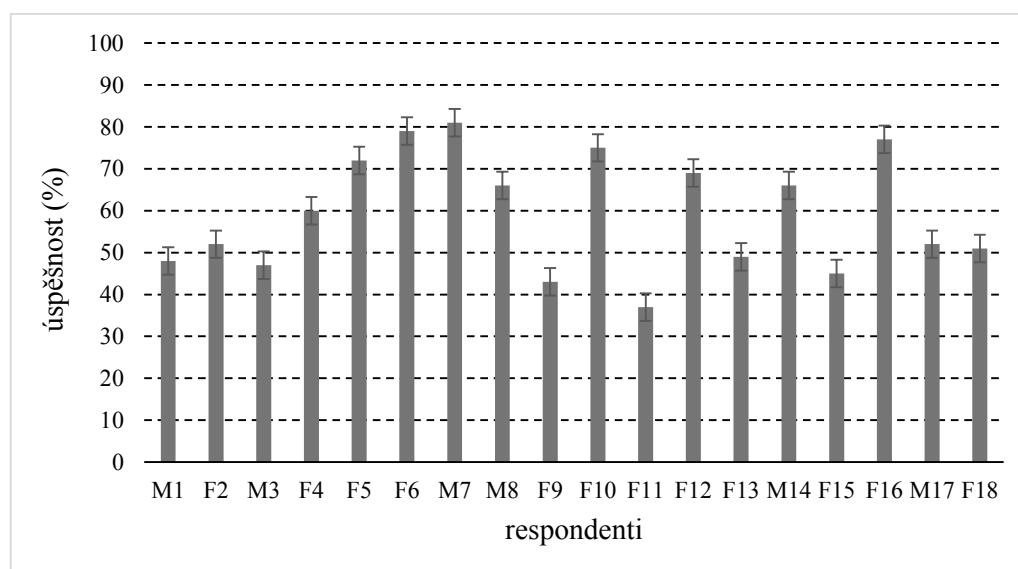
Obrázek 4.6 – Úspěšnost v položkách s manipulací temporální (T S) a v položkách s manipulací temporální i melodickou (T R).



Obrázek 4.7 – Úspěšnost podle typu položek a kroku manipulace: HH (hudební položky), RR (řečové položky), RH (smíšené položky), 1 (manipulace s krokem 1 ST), 0,5 (manipulace s krokem 0,5 ST).

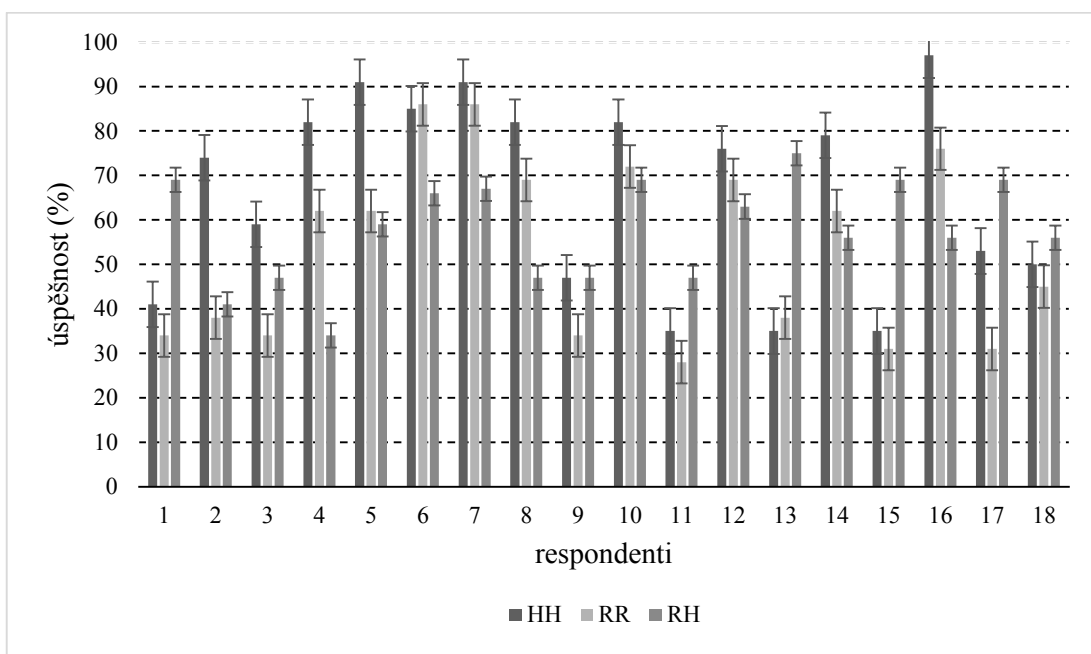
Z obrázku 4.5 je zřejmé, že položky obsahující manipulaci s krokem 1 ST mají prokazatelně vyšší úspěšnost než položky s manipulačním krokem 0,5 ST. Prohlédneme-li si graf na obrázku 4.7, který porovnává úspěšnost jednotlivých typů položek s krokem manipulace, zjistíme, že ne ve všech typech položek je výše zmíněné tvrzení tak jednoznačné. V položkách, které jsou minimálně z poloviny tvořené hudebními stimuly, tedy v položkách HH a RH, se chybové úsečky obou manipulačních kroků překrývají. V čistě řečových položkách (RR) se výše manipulačního kroku naopak projeví velmi zřetelně.

Graf na obrázku 4.8 přináší výsledky každého respondenta zvlášť. Žádné evidentní extrémy nejsou na první pohled z grafu patrné. Podíváme se tedy na úspěšnost jednotlivých respondentů podle typu položek.

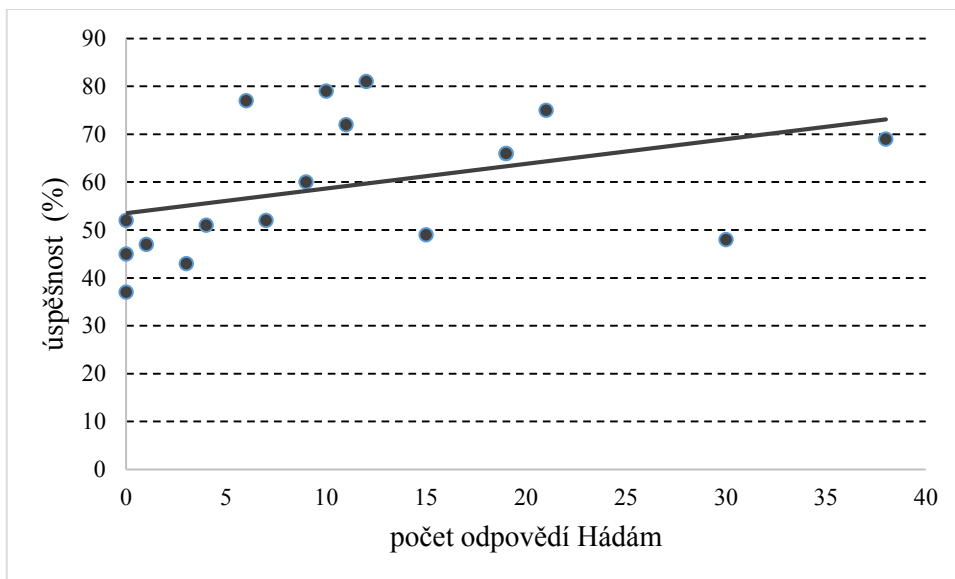


Obrázek 4.8 – Celková úspěšnost jednotlivých respondentů. Kód každého respondenta obsahuje jeho pohlaví (M = muži, F = ženy) a pořadí, ve kterém se testu účastnili.

V grafu na obrázku 4.9 vidíme, že někteří respondenti byli signifikantně méně či více úspěšní v různých typech položek. Můžeme si všimnout, že zatímco úspěšnost hudebních (HH), či smíšených (RH) položek u několika respondentů statisticky převyšuje ostatní typy, řečové položky (RR) jsou výše pouze u jediného respondenta, a to téměř zanedbatelně. Je patrné, že vysoká úspěšnost v jednom typu položek nezaručuje stejnou úspěšnost i v jiných typech.



Obrázek 4.9 – Úspěšnost jednotlivých respondentů v hudebních (HH), řečových (RR), a smíšených (RH) položkách.

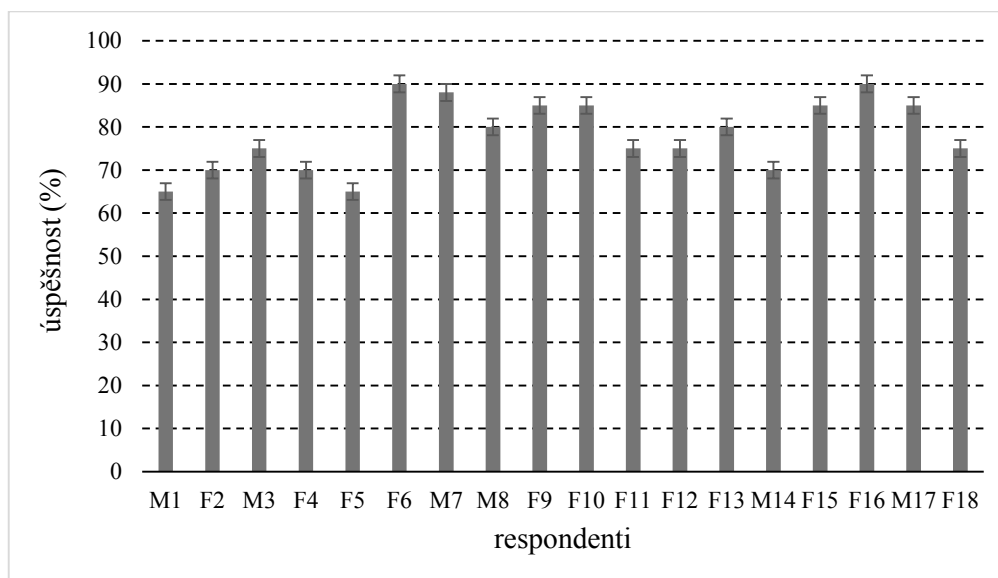


Obrázek 4.10 – Korelace úspěšnosti jednotlivých respondentů s počtem zaškrtnutých odpovědí Hádám.

Respondenti měli také možnost v záznamovém archu doplnit svou odpověď označením pole *Hádám*. Z dat jsme vypožorovali, že by její využití mohlo mít jistou souvislost s úspěšností jednotlivých respondentů. Graf na obrázku 4.10 znázorňuje

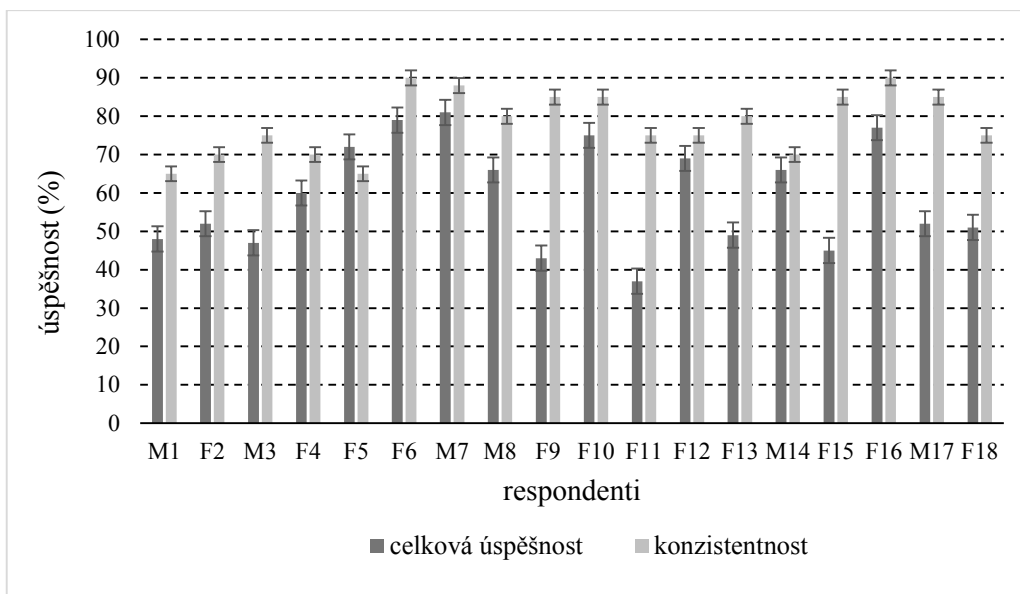
regresní přímku Pearsonova korelačního koeficientu, jehož hodnota se rovná 0,4. Takovou korelaci můžeme považovat za středně silnou.

Dále se podíváme na konzistentnost odpovědí všech respondentů. Připomeňme, že v celém testu bylo 20 opakovaných položek (z celkového počtu 95). U každého respondenta jsme spočítali, kolikrát se shodoval v odpovědích na tyto opakované otázky, a převedli výsledek na procenta (neboť u jednoho z respondentů nám chybělo několik hodnot). Vnitřní konzistentnost odpovědí každého respondenta je vynesena do grafu na obrázku 4.11 Celkově respondenti vykazují poměrně vysokou konzistentnost (78%). Ani v jednotlivých hodnotách nenalézáme vyloženě nekonzistentní odpovědi. Nejvyšší hodnoty dosahují 90%, nejnižší 65%. Zajímavé by bylo zjistit, zda nekoreluje vnitřní konzistentnost s celkovou úspěšností. Pro názornost jsme tedy vytvořili ještě jeden graf (viz obrázek 4.12). Kromě výsledků jednoho respondenta platí, že konzistentnost má vždy vyšší hodnoty, než celková úspěšnost. Ovšem přímou souvislost mezi těmito dvěma dimenzemi nepozorujeme. I méně úspěšní respondenti vykazují jedny z nejvyšších hodnot konzistentnosti.



Obrázek 4.11 - Vnitřní konzistentnost jednotlivých respondentů.





Obrázek 4.12 – Celková úspěšnost a vnitřní konzistentnost respondentů.

Podíváme-li se na nejméně úspěšné položky testu, zjistíme, že prvních dvanáct je typu S (stejně), z nichž úplně první je hudební (HH), a následující čtyři řečové (RR). U prvních čtyř dosahovala úspěšnost 100%, pátá 94%. V tabulce 4.13 si lze prohlédnout, z jakých slov byly položky vytvořeny. Nejméně úspěšné položky byly správně zodpovězeny pouze jedním, příp. třemi mluvčími, jejich úspěšnost je tak jen 6 nebo 17%.

nejúspěšnější položky				nejméně úspěšné položky			
typ	slovo	změna	úspěšnost	typ	slovo	změna	úspěšnost
HH	rána	S	100%	RR	řada	-0,5	6%
RR	jedna	S	100%	RR	chápe	+0,5	6%
RR	která	S	100%	RR	jaké	S T	6%
RR	kraje	S	100%	RH	chápe	+1	17%
RR	řeka	S	94%	RR	chápe	+0,5 T	17%

Tabulka 4.13 – Nejvíce a nejméně úspěšné položky testu. Typ položky: hudební (HH), řečová (RR), změna: stejné stimuly (S), -0,5 (manipulace -0,5 ST), +0,5 (manipulace +0,5 ST), S T (pouze temporální manipulace), +1 (manipulace +1 ST), +0,5 T (manipulace +0,5 ST i s temporální manipulací).

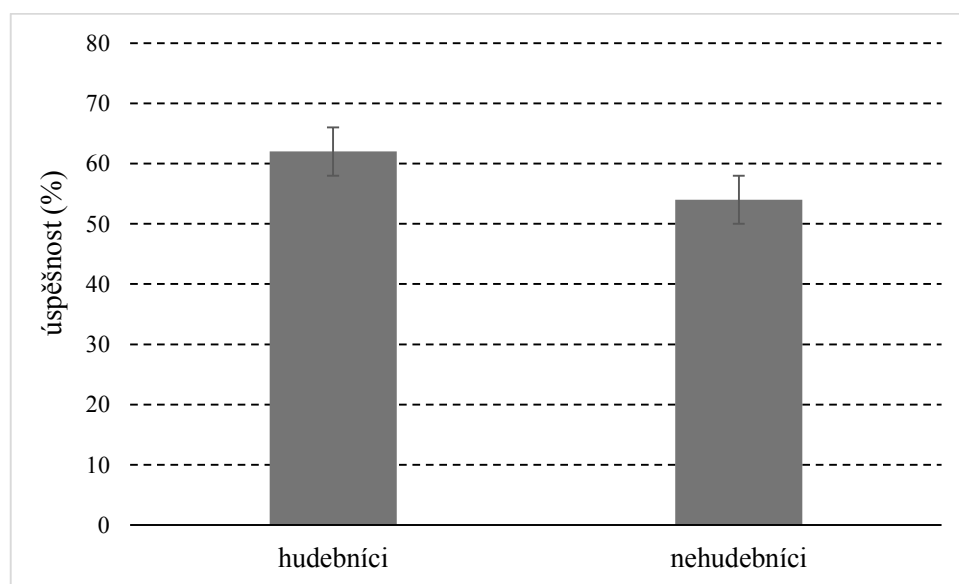
Nejméně úspěšné položky jsou také z většiny tvořeny řečovými položkami, pouze čtvrtá s nejhorší úspěšností je typu smíšeného. Ta jako jediná obsahuje

manipulaci +1 ST. V ostatních nejdeme změnu 0,5 ST oběma směry, příp. i s temporální manipulací. Ve třetí nejméně úspěšné položce je pouze temporální manipulace. Zatímco nejúspěšnější položky jsou zastoupeny odlišnými slovy, v nejméně úspěšných položkách najdeme jedno slovo (*chápe*) hned třikrát.

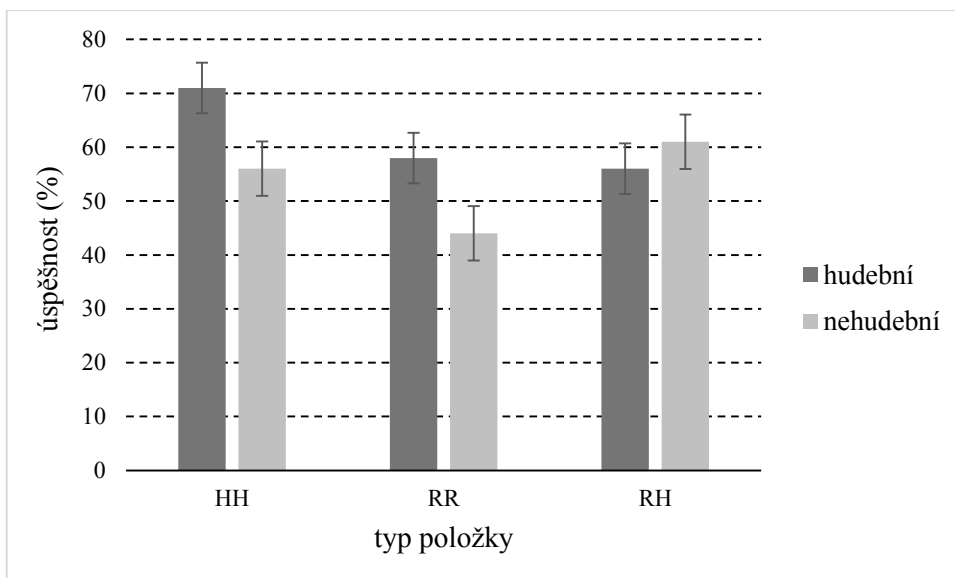
## 4.2 Hudebnost

Otázkám v dotazníku, který zjišťoval respondentovu hudebnost, byly přiřazeny body na škále 0-3 (ve druhé otázce 0,5-2,5). Při pěti otázkách byl tedy maximální počet 14,5 bodů. Respondenti byli poté dle hranice 7,25 bodů rozděleni na dvě skupiny – hudební a nehudební. Do hudební skupiny spadalo dvanáct respondentů, zbylých šest pak do skupiny nehudební.

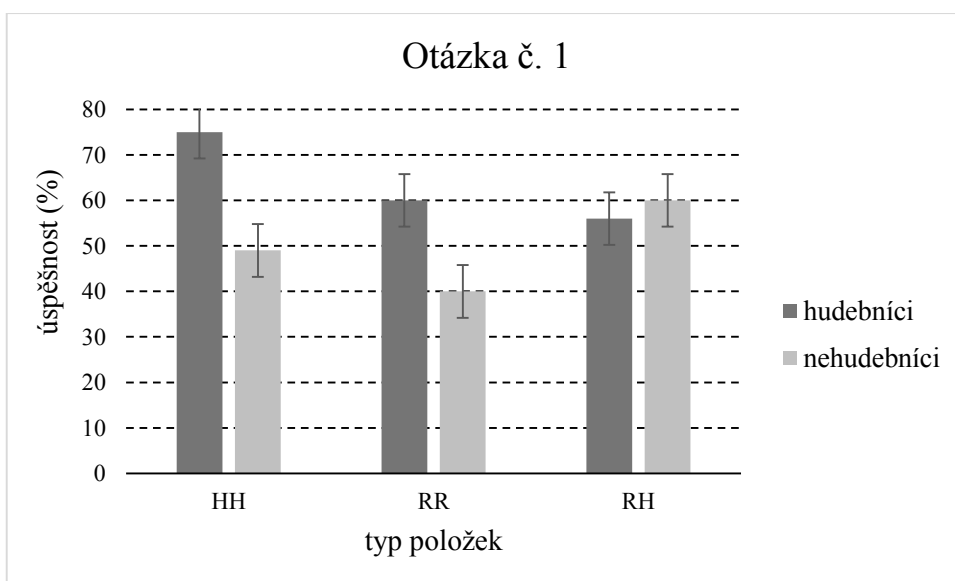
„Hudební“ respondenti byli celkově úspěšnější než „nehudební“. Z obrázku 4.14 můžeme vyčíst, že hudební respondenti dosahují úspěšnosti 62%, nehudební 54%. Podobné výsledky pozorujeme i v grafu na obrázku 4.15. Hudebníci jsou významně úspěšnější v položkách hudebních (HH) a řečových (RR). Ve smíšených položkách (RH) jsou naopak o něco úspěšnější nehudebníci.



Obrázek 4.14 – Celková úspěšnost hudebních a nehudebních respondentů.



Obrázek 4.15 – Úspěšnost hudebních a nehudebních respondentů podle typu položek.



Obrázek 4.16 – Úspěšnost hudebních a nehudebních respondentů podle otázky č. 1: Hudební vzdělání<sup>4</sup> v jednotlivých typech položek.

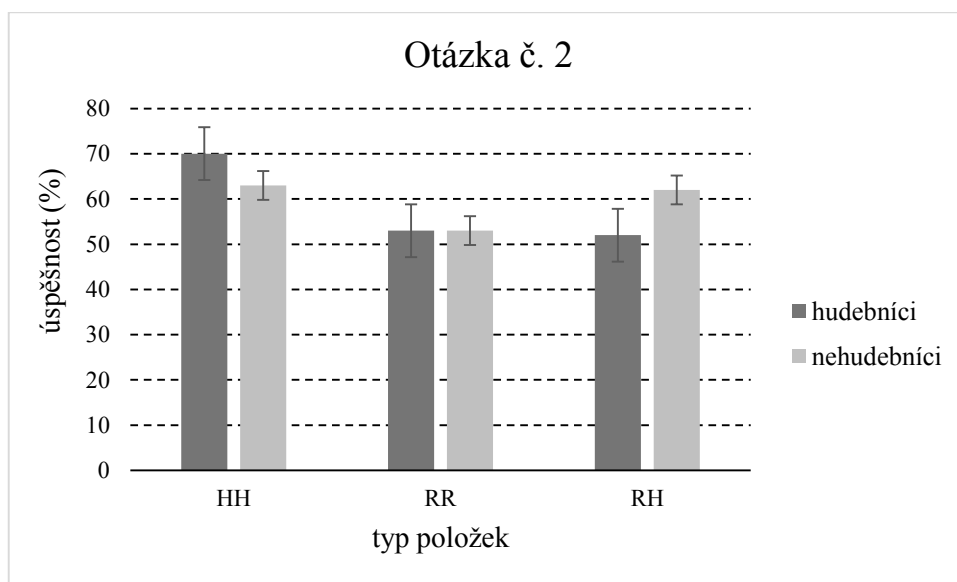
Nyní se podíváme, jaký mají na úspěšnost vliv jednotlivé otázky dotazníku. Maximální možný počet bodů každé z pěti otázek byl rozdělen na dvě poloviny (např. 0 - 1,5 a 2 - 3 body), a podle toho byli respondenti nově rozděleni na hudební a nehudební pro každou otázku zvlášť. Jeden respondent tak mohl být např. v jedné

<sup>4</sup> Podrobnosti o otázkách a jejich hodnocení v Příloze č. 2.

otázce hudebníkem a ve zbylých čtyřech nehudebníkem. Grafy na obrázcích 4.16 – 4.20 znázorňují výsledky.

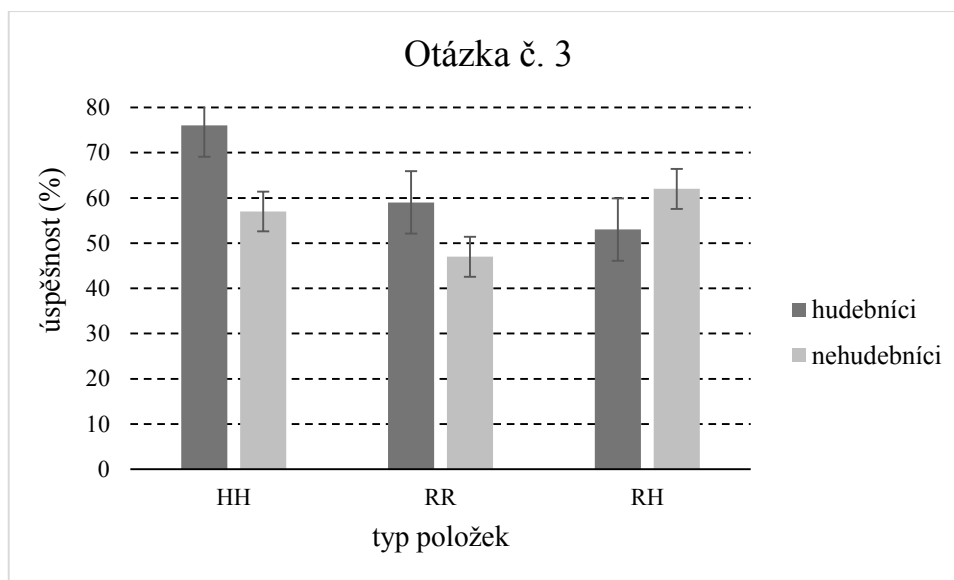
Podle otázky č. 1 jsme skupinu respondentů rozdělili na 12 hudebníků a 6 nehudebníků. Trend, který jsme viděli v grafu s úspěšností hudebních a nehudebních respondentů podle typu položek (viz obrázek 4.15), se v otázce č. 1 projevuje ještě zřetelněji (viz obrázek 4.16). Rozdíl mezi úspěšností nehudebníků (resp. subjektů s žádným nebo nízkým hudebním vzděláním) a hudebníků (resp. subjektů s minimálně ZUŠ vzděláním) v hudebních a řečových položkách je zde ještě o něco větší. Vyšší úspěšnost nehudebníků ve smíšených položkách je v případě otázky č. 1 velmi podobná jejich celkové úspěšnosti v tomto typu položek.

Výsledky osmi hudebníků a deseti nehudebníků rozdělených podle otázky č. 2: „Působili jste / působíte v nějakém hudebním tělese?“ se mírně odlišují od výsledků předchozí otázky. Vyšší úspěšnost hudebníků v hudebních položkách již není signifikantní (viz obrázek 4.17). V řečových položkách je dokonce stejná jako u nehudebníků. Ovšem poprvé pozorujeme významný rozdíl v položkách smíšených.



Obrázek 4.17 – Úspěšnost hudebních a nehudebních respondentů podle otázky č. 2: Působení v hudebním tělese.

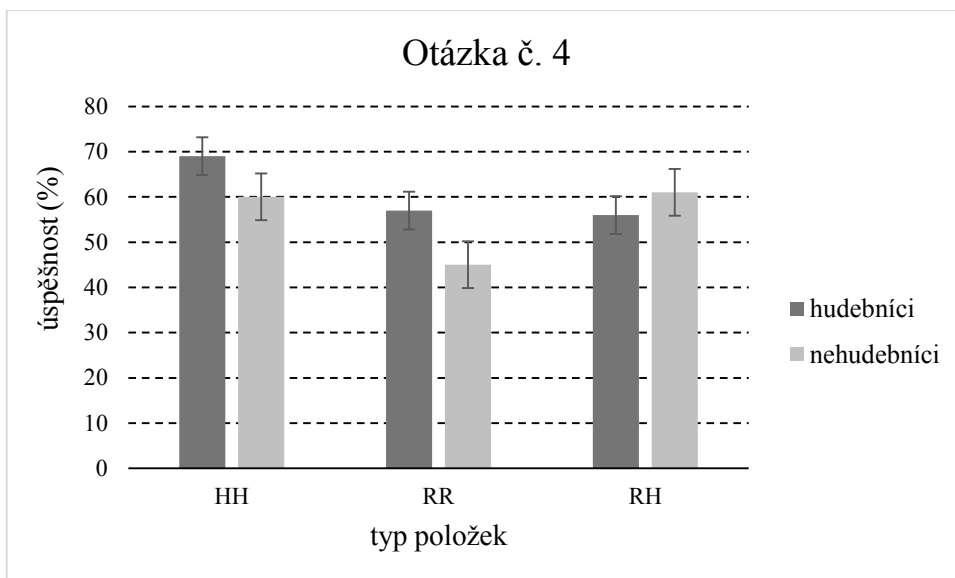
V otázce č. 3, která zjišťovala dobu hraní na hudební nástroj, se opět projevuje významný rozdíl mezi úspěšností obou skupin (zde rovnoměrně zastoupených po devíti respondentech) v hudebních a řečových položkách (viz obrázek 4.18). Signifikantní rozdíl zde ovšem (stejně jako v otázce č. 1) postrádá úspěšnost ve smíšených položkách.



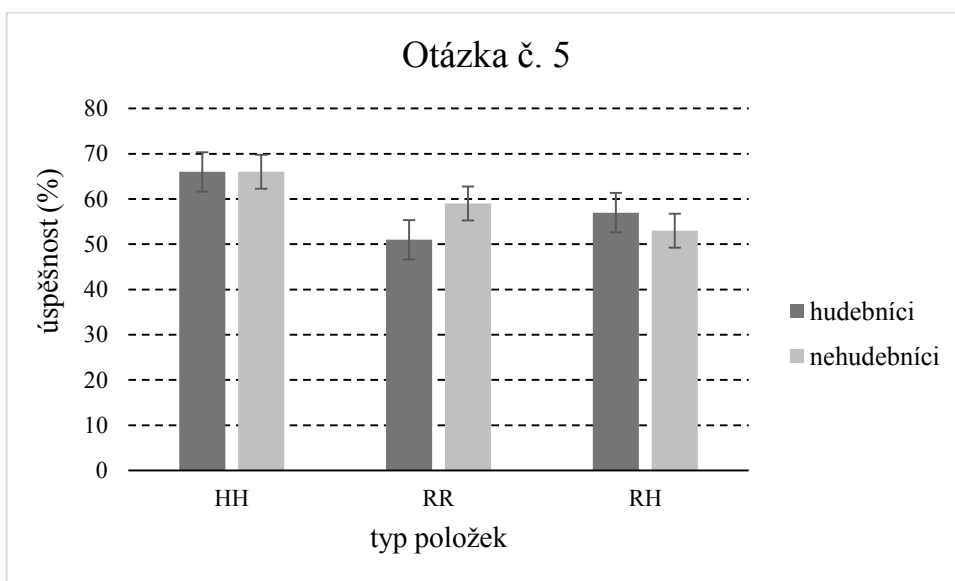
Obrázek 4.18 – Úspěšnost hudebních a nehudebních respondentů podle otázky č. 3: Doba hraní na nástroj/zpěvu v jednotlivých typech položek.

Otázka č. 4 se týkala četnosti cvičení na nástroj. Zde jsou výsledky pro dvanáct hudebníků (cvičících minimálně jednou týdně) a šesti nehudebníků (cvičícím maximálně jednou za měsíc nebo vůbec) velmi podobné výsledkům z předchozí třetí otázky. V grafu na obrázku 4.19 je můžeme vidět. Významné rozdíly zůstávají u hudebních a řečových položek, u smíšených položek vidíme dokonce o něco nižší rozdíl.

Graf na obrázku 4.20 přináší výsledky pro hudebníky a nehudebníky rozdělených dle vztahu k tanci (opět 12 hudebníků a 6 nehudebníků). Úspěšnost v hudebních položkách je v tomto případě shodná. Ovšem v řečových položkách vykazují „netanečníci“ (respondenti, kteří tančí jen příležitostně, či se tanci dokonce vyhýbají) vyšší úspěšnost než „tanečníci“ (respondenti, kteří mají tanec rádi, či ho dokonce provozují profesionálně). Ve smíšených položkách jsou poprvé o něco úspěšnější hudebníci.



Obrázek 4.19 – Úspěšnost hudebních a nehudebních respondentů podle otázky č. 4: Četnost cvičení na nástroj v jednotlivých typech položek.



Obrázek 4.20 – Úspěšnost hudebních a nehudebních respondentů podle otázky č. 5: Vztah k tanci v jednotlivých typech položek.

## 5. Diskuse

### 5.1 Úspěšnost

Celková úspěšnost v rozpoznávání melodických a temporálních manipulací byla naměřena 59,4%. Tato poněkud nízká hodnota značí vysokou obtížnost testu. Nízkou úspěšnost způsobuje také fakt, že v testu převažovaly manipulované položky nad nemanipulovanými. Pokud by tomu bylo naopak, byla by úspěšnost i u posluchačů, kteří slyšeli rozdíly pouze zřídka, vyšší. Nejspíše proto položky typu S (stejně) vykazují vyšší úspěšnost (81%) než položky typu R (různé; 50%). Ze stejného důvodu má také nejvyšší úspěšnost blok č. 1, který obsahoval nejvíce položek typu S, a to hlavně v případě, že byl zadáván jako první v pořadí. Posluchači se mohli během prvních položek bloku č. 1 adaptovat na výši manipulačních kroků, jejichž rozpoznání test vyžadoval. Naopak nejnižší úspěšnost bloku č. 2 připisujeme tomu, že na něj připadlo nejvíce položek s nejnižším manipulačním krokem 0,5 ST.

Nejvyšší úspěšnost hudebních položek jsme předpokládali. Jak jsme již v kapitole 2 na str. 18 a 19 uvedli, nejmenší vnímatelné rozdíly se v řečovém stimulu navyšují. Rozpoznání melodických, příp. temporálních, změn bylo tak pro naše posluchače jednodušší u hudebních položek. Jakmile položka obsahovala alespoň jeden stimul řečový, stala se pro posluchače percepčně náročnější. U položek čistě řečových byla úspěšnost očekávatelně nejnižší. Úspěšnost RR položek byla dokonce tak nízká (48%), že zasahuje pod průměrnou úroveň náhody. Tento pozoruhodný jev je jistě ukazatelem nízkého počtu respondentů.

Výsledky podle typu manipulace ukazují, že položky s melodickou změnou 1 ST byly významně úspěšnější než položky se změnou 0,5 ST, či dokonce s temporální manipulací. Hodnoty položek s temporální či melodickou manipulací +0,5 ST a -0,5 ST zasahovaly opět pod úroveň hranice náhody. Snazší rozpoznatelnost větších melodických změn je předvídatelná, ovšem překvapivá je nízká úspěšnost položek s temporální změnou (29%). A to i přesto, že v grafu na obrázku 4.5 jsou zahrnuty výsledky položek, které obsahovaly pouze temporální nebo temporální i melodickou manipulaci. Položky s oběma manipulacemi dokonce převažovaly. Výsledky pro položky s temporální manipulací ukazují, že pouze

temporální změna 25 ms v jinak totožných stimulech byla prakticky nepostřehnutelná. Úspěšnost položek s oběma typy manipulace dosahuje alespoň 51%. Připomeňme ovšem, že temporální manipulace sloužila v našem testu hlavně jako distraktor. Pro konkrétní závěry ohledně temporálně manipulovaných položek by bylo třeba dalších experimentů.

Naše výsledky také nekorrespondují s hodnotami autorů, které jsme citovali v kapitole 2. Např. Fastl, Zwicker (2007) uvádějí, že rozdíl půl půltónu je i v řeči postřehnutelný. Podle výsledků z našeho experimentu však tento rozdíl v řeči postřehnutelný není. Z grafu na obrázku 4.7, který jsme uvedli v předchozí kapitole, je zřejmé, že rozdíl půl půltónu je rozpoznatelný sotva v hudebních či smíšených položkách. V řečových položkách jsou manipulační kroky 0,5 i 1 ST pod úrovní 50% úspěšnosti. Přestože výsledky opět zkresluje malý vzorek subjektů, výše zmíněné tendence jsou natolik extrémní, že není vyloučené, že by se s více respondenty nepotvrdily. Řečové položky jsou také jediné, ve kterých je úspěšnost obou typů manipulačních kroků významně odlišná. Zdá se, že v řeči je každé snížení manipulačního kroku pro jeho percepci radikální.

Z výsledků pro jednotlivé respondenty dle typu položek je patrné, že pokud byl jeden respondent úspěšný v jistém typu položek, nemusel být stejně úspěšný v typech ostatních. Často znatelně převyšovala úspěšnost v hudebních, příp. smíšených položkách. Nejvyšší úspěšnost v řečových položkách pozorujeme pouze v jediném případě, a to ještě bez statistické významnosti. Můžeme tedy potvrdit výše zmíněné tendence pro lepší percepci melodie v hudebních než v řečových položkách. Někteřím respondentům na úspěšnosti ubírala schopnost (resp. neschopnost) překonat odlišnosti řeči a hudby, a soustředit se tak na změny ve smíšených položkách. Takových respondentů byla přibližně polovina. Pozoruhodné jsou pak výsledky respondentů, kteří vykazovali poměrně nízkou úspěšnost v hudebních a řečových položkách, a ve smíšených naopak úspěšnost velmi vysokou. Tito respondenti tvořili také téměř polovinu našeho vzorku. Tuto nenadálou úspěšnost lze opět připisovat vyššímu počtu položek typu *Různé* oproti položkám *Stejným*. Je možné, že jinak „neúspěšní“ respondenti hodnotili smíšené položky jako různé už jen z důvodu odlišnosti jejich modality. Podpořením naší domněnky může být fakt, že až na jedinou výjimku je polovina našich celkově



úspěšnějších respondentů nejúspěšnější v položkách hudebních, pak v řečových a až na posledním místě ve smíšených.

Případnou nejistotu ve své odpovědi mohli respondenti vyznačit zaškrtnutím pole *Hádám*. Podívali jsme se, zda míra nejistoty souvisí s úspěšností jednotlivých respondentů či nikoliv. Korelační Pearsonův koeficient 0,4 značí středně silnou korelaci. V našem případě bychom očekávali korelaci zápornou, tedy čím vyšší nejistota, tím nižší úspěšnost respondenta. Ovšem výsledky naznačují opak. Respondenti, kteří častěji označili možnost *Hádám*, vykazují vyšší úspěšnost. Jev může být způsoben náročností percepčního testu. Některé melodické či temporální změny byly pro percepčně nadanější posluchače zaznamenatelné, ovšem ne natolik, aby si v tom byli jisti. Korelace může být také jen důsledkem vyšší zodpovědnosti úspěšnějších respondentů v jejich zaznamenávání.

Vnitřní konzistentnost našich respondentů činí 78%. To je poměrně uspokojivé číslo, na jehož základě můžeme naše domněnky uvádět s větší jistotou. Nezdá se ale, že by konzistentnost úzce souvisela s celkovou úspěšností. I méně úspěšní respondenti byli ve svých odpovědích velmi konzistentní.

Položková analýza ukázala, že nejúspěšnější položky byly typu S, tedy takové, které obsahovaly stejné stimuly. Opět se zde mohla projevit problematika neslyšených manipulací způsobující automatický nárůst úspěšnosti v položkách stejných. Ovšem čtyři z pěti celkově nejúspěšnějších položek vykazují 100% úspěšnost, tudíž pokládáme tento efekt za zanedbatelný. Vysoká úspěšnost ve stejných položkách je také uspokojujícím ukazatelem kvality našeho testu. Druhá, třetí a čtvrtá nejúspěšnější položka spadá pod řečový typ. Může se zdát, že řečové položky zde najednou dosahují své pozbývané úspěšnosti, ovšem to, že byly hodnoceny (sice správně) jako stejné, nevylučuje možnost, že jako stejné byly označovány ve většině případů. Obtížnosti řečového signálu se plně projeví až s výší manipulačního kroku. Ve stejném počtu (čtyři z pěti) se řečové položky objevují i mezi nejméně úspěšnými položkami. Zde již vidíme, že manipulační kroky 0,5 ST jsou nejhůře rozpoznatelné. Objevuje se dokonce i položka s manipulací melodickou + 0,5 ST i temporální. Temporální manipulaci pozorujeme ještě u třetí nejméně úspěšné položky, opět v souvislosti s řečí. Jedinou neřečovou položkou v této neúspěšné skupině je smíšená položka pojící se s manipulačním krokem +1 ST a slovem *chápe*. V obtížnosti této položky může

hrát roli původní interval mezi vokály, základní frekvence hlasu mluvčího, směr melodie, dokonce i typ vokálů. Ačkoliv jsme pro náš experiment vybrali dvojslabičná slova pouze s vokály /a/ a /e/, není vyloučené, že jejich kvalita neovlivňuje percepci melodie. Řešení této otázky by přesahovalo možnosti této práce. Důležitým poznatkem z položkové analýzy je ten, že nejúspěšnější položky jsou typu S, a stejně jako ty nejméně úspěšné obsahují převážně řečové položky. V nejméně úspěšných položkách pozorujeme temporální či nejmenší melodické manipulace. Nenajdeme v nich ani jednu položku čistě hudební.

## 5.2 Hudebnost

Hudební respondenti byli o něco úspěšnější (62%) než nehudební (54%). Zdá se, že hudebnost jim pomohla v hudebních i v řečových položkách, tam dokonce ještě zřetelněji. Nelze ovšem na základě našich výsledků tvrdit, že by hudebně nadaní lidé využívali svou hudebnost více v řeči než v hudbě. Obecně můžeme shrnout, že hudebně citlivější posluchači mají v percepci melodie v řeči a hudbě výhodu. Jejich výhoda se ale neprokázala u smíšených položek. O něco vyšší úspěšnost nehudebníků ve smíšených položkách můžeme připisovat již zmíněnému efektu odlišnosti modalit, nebo se hudebnost našich respondentů omezuje pouze na homogenní položky. Jelikož je úspěšnost nehudebníků v těchto položkách nečekaně vyšší než v hudebních (o 5%) a již statisticky významně v řečových (o 17%) položkách, přikládáme se spíše k první z možných příčin.

Některé výsledky vlivu hudebnosti podle jednotlivých otázek jsou podobné výsledkům celkovým. Úspěšnost hudebníků a nehudebníků podle otázek č. 1 (Hudební vzdělání) a č. 3 (Doba hraní na nástroj/zpěv) se s celkovou úspěšností shodují. Výsledky otázky č. 4 (Četnost cvičení na nástroj/zpěv) se liší pouze v nižší signifikanci převahy hudebníků nad nehudebníky v HH položkách. Můžeme tedy říci, že pro respondenty hudebně vzdělanější, kteří jsou s nástrojem/zpěvem v kontaktu častěji a déle, platí výše zmíněné tendence, tedy zřetelná výhoda při vnímání melodie v řeči a v hudbě. V otázce č. 2 (Působení v hudebním tělese) je vyšší úspěšnost hudebníků v HH položkách již spíše nesignifikantní. Ovšem stále

můžeme výhodu hudebně aktivních při vnímání melodie v hudbě předpokládat. V řečových položkách je úspěšnost hudebníků podle otázky č. 2 dokonce stejná jako úspěšnost nehudebníků. Zdá se tedy, že dle našich výsledků nezaručuje posluchači působení v hudebním tělese přínos pro percepci melodie v řeči. Vyšší úspěšnosti nehudebníků dle otázky č. 2 nepřipisujeme žádné nové příčiny. Otázka č. 5 (Vztah k tanci) přinesla zajímavé výsledky. Úspěšnost tanečníků a netanečníků v hudebních položkách je totožná. Můžeme usuzovat, že vztah k tanci nemá vliv na vnímání melodie v hudbě. Případně by šlo uvažovat o tanci jako nevhodném ukazateli hudebnosti. V řečových položkách jsou překvapivě úspěšnější naopak netanečníci. Jejich znatelně vyšší úspěšnost mohla způsobit opět skutečnost, že kladný vztah k tanci nekoreluje s hudebností jako takovou. V netaneční skupině se tak mohli objevit i jinak hudební respondenti. Podíváme-li se do dat, zjistíme, že tomu tak opravdu z poloviny je. O něco vyšší úspěšnost tanečníků ve smíšených položkách připisujeme stejnému jevu. Tanec tedy dle našich dat nezaručuje hudebnost, ani percepční výhody, které z ní vycházejí. Musíme také vzít v úvahu, že odpovědi na pátou otázku dotazníku jsou do jisté míry subjektivní, vyjadřují pouze vztah k tanci, čili nikoliv úspěšnost v něm.

### **5.3 Limity studie a další výzkum**

Nejzásadnějším omezením naší studie se ukázal být nízký počet respondentů. Navazující výzkum s vyšším počtem subjektů by mohl potvrdit či zpřesnit naše výsledky, nebo dokonce odhalit další zajímavé tendence. Kupříkladu by bylo zajímavé zjistit, zda nemá na výsledky vliv velikost původního intervalu mezi vokály a směr jeho změny. V našem experimentu jsme na tuto skutečnost nebrali ohled. Podle nás ale může hrát roli, zda je výchozím intervalem tercie či septima, a zda je percepčně snadnější postřehnout její snížení či zvýšení. Ostatně vlivem směru melodické změny v hudbě a řeči se zabývali Weidema, Roncaglia-Denissem a Honing (2016), jejichž experiment v kapitole 2.3 zmiňujeme. Zároveň by bylo zajímavé porovnat výsledky s manipulacemi realizovanými na druhé slabice, jak jsme postupovali my, s výsledky manipulací na slabice první. Dle našich pilotních respondentů by takové výsledky nedosahovaly příliš vysokých hodnot. V našem experimentu jsme se také věnovali zejména melodii, další výzkum by se mohl

podrobněji zaměřit na temporální složku. Výsledky také mohly být ovlivněny typem vokálu. Ačkoliv jsme se omezili pouze na dvě české samohlásky, není vyloučen vliv jejich kvality. Významnější roli by také mohli hrát mluvčí a hodnota jejich základní hlasové frekvence. Náš experiment by se také dal vylepšit vhodnějším zadáváním percepčního testu a jeho zaznamenáváním. Respondenti si často nepamatovali všechny informace, které jim ohledně testu byly podány. Snad by individuálnější a důraznější, i když jistě časově náročnější, přístup napomohl lepšímu pochopení všech instrukcí. Záznamový arch by mohl být členitější, a tím graficky přehlednější. Někteří z respondentů se v souvislém sloupci třiceti dvou úzkých řádků ztráceli (viz Příloha č. 1). Doporučili bychom sloupce rozdělit na více částí, příp. vložit určitý předěl i do zvukové stopy. Další výzkum by se také mohl zevrubně věnovat hudebnosti. Bylo by zajímavé zjistit, jaké faktory přesně utváří onu hudebnost, a které z nich jsou v percepci melodie stěžejní. Samozřejmě by bylo přínosem výzkum rozšířit na více věkových skupin, a prozkoumat, jak se percepce melodie v řeči a hudbě v různém věku či pohlaví liší.

## 6. Závěr

V této práci jsme se zaměřili na percepci melodie v řeči a v hudbě. Ve druhé kapitole jsme nastínili teoretické pozadí problematiky a uvedli stěžejní současné studie zabývající se melodií. Třetí, čtvrtá a pátá kapitola se věnovala experimentu, který jsme na toto téma vytvořili. Sestavili jsme percepční test, který zjišťoval jaké melodické (příp. temporální) kroky jsou v hudbě a v řeči postřehnutelné. Výsledky našeho experimentu ukázaly, že v hudbě jsou melodické manipulace obecně snáze rozpoznatelnější než v řeči, a to zejména s manipulačními kroky 1 ST. Temporální a menší melodické kroky se navzdory poznatkům z uváděné literatury ukázaly být percepčně spíše nepostřehnutelné. Zaměřili jsme se také na hudebnost a její vliv na percepci melodie. Ukázalo se, že respondenti označení jako „více hudební“ své dovednosti využívají ve vnímání melodie v řeči i v hudbě.

Přínosem této práce je seznámení čtenáře se současnými poznatky problematiky melodie a uvedení do teoretického zázemí postřehnutelných rozdílů. Experiment, který jsme provedli, je navazujícím článkem v rámci tohoto tématu. Validitu výsledků našeho experimentu podporuje důraz na přirozená data, ze kterých jsme vycházeli. Doufáme, že navazující výzkum bude pokračovat stejným směrem, a potvrdí tak naše domněnky.

## Bibliografie

- BACHEM, A., 1937. *Various types of absolute pitch*. Journal of the Acoustical Society of America, 9, s. 146-151.
- BARLOW, H., MORGENSTERN, S., 1983. *A Dictionary of Musical Themes*. In: PATEL, A. D., IVERSEN, J. R., ROSENBERG, J. C., 2006. *Comparing the rhythm and melody of speech and music: The case of British English and French*. Journal of the Acoustical Society of America, 119.
- BOERSMA, P., 1993. *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound*. in Proceedings of the Institute Of Phonetic Sciences, University of Amsterdam, 17, s. 97–110.
- BOERSMA, P., WEENINK, D., 2014. Praat: doing phonetics by computer [počítačový program]. v5.3.73 [online]. 2014. Dostupné na: <http://www.praat.org/>
- BOŘIL, T., 2016. *Psychoakustika* [online]. Dostupné na: [http://fonetika.ff.cuni.cz/wp-content/uploads/sites/104/2016/06/3\\_psychoakustika.pdf](http://fonetika.ff.cuni.cz/wp-content/uploads/sites/104/2016/06/3_psychoakustika.pdf)
- CLARK, J., YALLOP, C., 1995. *An Introduction to Phonetics and Phonology. Second Edition*. Massachusetts: Blackwell Publishing, s. 306-308, 322-333.
- CLARK, J., YALLOP, C., FLETCHER, J., 2007. *An Introduction to Phonetics and Phonology. Third Edition*. Oxford: Blackwell Publishing, s. 308, 332.
- COOPER, W. E., SORENSEN, J. M., 1981. *Fundamental frequency in sentence production*. New York/Heidelberg/Berlin: Springer.
- CRUTTENDEN, A., 1997. *Intonation. Second Edition*. New York: Cambridge University Press, s. 3.
- FASTL, H., ZWICKER, E., 2007. *Psychoacoustics. Facts and Models*. Heidelberg: Springer, s. 113, 176-180, 183-186, 265-274.
- FERNÁNDEZOVÁ, E. M., SMITHOVÁ-CAIRNSOVÁ, H., 2014. *Základy psycholingvistiky*. Praha: Karolinum, s. 164

FISCHER-JORGENSEN, E., 1961. What Can the New Techniques of Acoustic Phonetics Contribute to Linguistics? In: SAPORTA, S. 1961. *Psycholinguistics. A book of readings*. New York: Holt, Rinehart, Winston, s. 132.

FLANAGAN, J. L., SASLOW, M. G., 1958. *Pitch discrimination for synthetic vowels*. Journal of the Acoustical Society of America, 30.

FOSS, D. J., HAKES, D. T., 1978. *Psycholinguistics. An introduction to the psychology of language*. New Jersey: Prentice-Hall, s. 61-62.

FRIČ, M., 2011. *Efekt zvyšování hlasitosti na spektrální charakteristiky hlasu u různých typů použití hlasu a u různých skupin hlasových profesí*. Akustické listy. roč. 17, č. 1-2, s. 19-25.

HARDCASTLE, W. J., LAVER, J., GIBBON, F., 2010. *The Handbook of phonetics sciences*. Oxford: Wiley-Blackwell s. 469-471.

HARRIS, M. S., UMEDA, N., 1986. *Difference limens for fundamental frequency contours in sentences*. Journal of the Acoustical Society of America, 81.

HOWARD, D., ANGUS, J., 2009. *Acoustics and Psychoacoustics. Fourth edition*. Focal Press, s. 97-98, 137-141, 242-243, 258.

ISSACHENKO, A. V., SCHADLICH, H.-J., 1970. *A model of standard German intonation*. In: T'HART, J., COLLIER, R., COHEN, A., 1990. *A Perceptual Study of Intonation. An experimental-phonetic approach to speech melody*. New York: Cambridge University Press, s. 28.

KLATT, D. H., 1973. *Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception*. Journal of the Acoustical Society of America, 53.

LEE, C. S., TODD, N. P. McA., 2004. *Toward an auditory account of speech rhythm: Application of a model of the auditory 'primal sketch' to two multi-language corpora*. Cognition, 93.

MAIDMENT, J., 1976. *Voice fundamental frequency characteristics as language*

*differentiators*. In: PATEL, A. D., IVERSEN, J. R., ROSENBERG, J. C., 2006. *Comparing the rhythm and melody of speech and music: The case of British English and French*. Journal of the Acoustical Society of America, 119.

MCSHEFFERTY, D., WHITMER, W. M., AKEROYD, M. A., 2015. *The just-noticeable difference in speech-to-noise ratio*. Trends in Hearing, 19. Dostupné na: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4335553/>

MERTENS, P., 2004. *The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model*. In: Proceedings of Speech Prosody 2004, Nara, Japan, s. 23–26.

MOORE, B. C. J., PETERS, R. W., GLASBERG, B. R., 1993. *Detection of temporal gaps in sinusoids: effects of frequency and level*. Journal of the Acoustical Society of America, 93.

MORTON, J., MARCUS, S. M., FRANKISH, C. R., 1976. *Perceptual centres (P-Centres)*. Psychological Review, 83.

NAZZI, T., BERTONCINI, J., MEHLER, J., 1998. *Language discrimination in newborns: Toward an understanding of the role of rhythm*. J. Exp. Psychol. Hum. Percept. Perform., 24.

PATEL, A. D., DANIELE, J. R., 2003. *An empirical comparison of rhythm in language and music*. Cognition 87.

PATEL, A. D., IVERSEN, J. R., ROSENBERG, J. C., 2006. *Comparing the rhythm and melody of speech and music: The case of British English and French*. Journal of the Acoustical Society of America, 119.

PIERREHUMBERT, J., 1979. *The perception of fundamental frequency declination*. Journal of the Acoustical Society of America, 66.

PISONI, D. B., 1977. *Identification and discrimination of the relative onset times of two component tones: implications for voicing perception in stops*. Journal of the Acoustical Society of America, 61.



POLLACK, I., 1968. *Detection of rate of change of auditory frequency*. Journal of Experimental Psychology, 11.

RAMUS, F., NESPOR, M., MEHLER, J., 1999. *Correlates of linguistic rhythm in the speech signal*. Cognition, 73.

RAMUS, F., 2002. *Acoustic correlates of linguistic rhythm: Perspectives*. In Proceedings of Speech Prosody, 2002, Aix-en-Provence, s. 115–120.

REPP, B. H., SCHOUTEN, M. E. H., 1987. *The Psychophysics of Speech Perception*. Utrecht: Martinus Nijhoff Publishers, s. 11, 273.

ROSSI, M. 1971. *Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole*. In: T'HART, J., COLLIER, R., COHEN, A., 1990. *A Perceptual Study of Intonation. An experimental-phonetic approach to speech melody*. New York: Cambridge University Press, s. 31.

ROSSI, M., CHAFCOULOFF, M., 1972. *Recherche sur le seuil différentiel de fréquence fondamentale dans la parole*. In: T'HART, J., COLLIER, R., COHEN, A., 1990. *A Perceptual Study of Intonation. An experimental-phonetic approach to speech melody*. New York: Cambridge University Press, s. 28.

ROSSI 1978. *La perception des glissandos descendants dans les contours prosodiques (The perception of falling glissandos in prosodic contours)*. In: T'HART, J., COLLIER, R., COHEN, A., 1990. *A Perceptual Study of Intonation. An experimental-phonetic approach to speech melody*. New York: Cambridge University Press, s. 31.

SERGEANT, R. L., HARRIS, J. D., 1962. *Sensitivity to unidirectional frequency modulation*. Journal of the Acoustical Society of America, 34.

SKARNITZL, R., ŠTURM, P., VOLÍN, J., 2015. *Zvuková báze řečové komunikace*. Praha: Karolinum, s. 37-43, 150.

STREETER, L., 1978. *Acoustic determinants of phrase boundary perception*. Journal of the Acoustical Society of America 64.

THART, J., COLLIER, R., COHEN, A., 1990. *A Perceptual Study of Intonation. An experimental-phonetic approach to speech melody*. New York: Cambridge University Press, s. 27-36.

VENCOVSKÝ, V., 2014. *Roughness Prediction for Complex Acoustic Stimuli*. *Akustické listy*. roč. 20, č. 3–4, s. 19–26

WEIDEMA, J.L., RONCAGLIA-DENISSEN, M. P., HONING H., 2016. *Top-Down Modulation on the Perception and Categorization of Identical Pitch Contours in Speech and Music*. *Frontiers in Psychology*, 7.

## **Přílohy**



## Příloha č. 2 – Hudební dotazník

KÓD: .....

Věk: .....

### 1) Hudební vzdělání:

ZUŠ - KONZERVATOŘ - VŠ/VOŠ - SOUKROMÉ LEKCE - SAMOUK – ŽÁDNÉ

2	2,5	3	1,5	1	0	(body)
---	-----	---	-----	---	---	--------

### 2) Působili jste / působíte v nějakém hudebním tělese? (*sbor, trio, kapela, orchestr...*)

ano	2,5
ne	0,5

### 3) Jak dlouho jste hráli / hrajete na hudební nástroj (resp. zpíváte)? (*údaj v letech*)

méně než 1	0
1 až 4	1
5 až 10	1,5
11 až 15	2
16 až 20	2,5
více než 20	3

### 4) Jak často jste cvičili / cvičíte na hudební nástroj (resp. zpěv)? (*např. 3x týdně, každý den, jednou za měsíc, sporadicky*)

vůbec	0
sporadicky	1
1x/měs.	1,5
1-x/týden	2
3x/týden-ob den	2,5
denně	3

### 5) Jaký máte vztah k tanci? (*vyhýbám se mu / nezajímá mě / rád tancuji, když můžu / tancuji amatérsky / profesionálně...*)

nikdy, nezájem	0
příležitostně	1
rád příležitostně/amatérsky	2
profesionálně	3

**celkem 0,5-14,5**